Functional Characterization of Nine Norway Spruce *TPS* Genes and Evolution of Gymnosperm Terpene Synthases of the *TPS-d* Subfamily[1][w]

Diane M. Martin, Jenny Fäldt, and Jörg Bohlmann*

Biotechnology Laboratory (D.M.M., J.F., J.B.), and Departments of Botany (D.M.M., J.B.) and Forest Sciences (J.B.), University of British Columbia, Vancouver V6T 1Z3, British Columbia, Canada

Constitutive and induced terpenoids are important defense compounds for many plants against potential herbivores and pathogens. In Norway spruce (*Picea abies* L. Karst), treatment with methyl jasmonate induces complex chemical and biochemical terpenoid defense responses associated with traumatic resin duct development in stems and volatile terpenoid emissions in needles. The cloning of (+)-3-carene synthase was the first step in characterizing this system at the molecular genetic level. Here we report the isolation and functional characterization of nine additional terpene synthase (*TPS*) cDNAs from Norway spruce. These cDNAs encode four monoterpene synthases, myrcene synthase, (−)-limonene synthase, (−)-α/β-pinene synthase, and (−)-linalool synthase; three sesquiterpene synthases, longifolene synthase, *E,E*-α-farnesene synthase, and *E*-α-bisabolene synthase; and two diterpene synthases, isopimara-7,15-diene synthase and levopimaradiene/abietadiene synthase, each with a unique product profile. To our knowledge, genes encoding isopimara-7,15-diene synthase and longifolene synthase have not been previously described, and this linalool synthase is the first described from a gymnosperm. These functionally diverse TPS account for much of the structural diversity of constitutive and methyl jasmonate-induced terpenoids in foliage, xylem, bark, and volatile emissions from needles of Norway spruce. Phylogenetic analyses based on the inclusion of these TPS into the TPS-d subfamily revealed that functional specialization of conifer TPS occurred before speciation of Pinaceae. Furthermore, based on TPS enclaves created by distinct branching patterns, the TPS-d subfamily is divided into three groups according to sequence similarities and functional assessment. Similarities of TPS evolution in angiosperms and modeling of TPS protein structures are discussed.

Terpenoids are the largest class of specialized plant natural products (Croteau et al., 2000), which form an essential part of direct and indirect defense systems against herbivores and pathogens. Terpenoids can function in direct defense as toxins, feeding and oviposition deterrents, or insect hormone analogs. Low M_r terpenoids are often released as volatiles from plants during attack by herbivores. These emissions can attract predators and parasites of arthropod herbivores in certain forms of indirect defense (Kessler and Baldwin, 2002). Two classes of terpenoids, the monoterpenoids (10 carbon atoms) and diterpenoids (20 carbon atoms), provide the majority of constitutive and induced oleoresin defense metabolites in conifers (Bohlmann and Croteau, 1999; Phillips and Croteau, 1999; Seybold et al., 2000; Trapp and Croteau, 2001a; Martin et al., 2002; Fig. 1). Resin terpenoids are sequestered in specialized anatomical structures, such as resin ducts or resin blisters. Upon damage of resin ducts by stem boring insects, diterpenoids solvated with monoterpenoids exude, creating a long-lasting chemical and physical barrier at the site of insect attack. Sesquiterpenoids (15 carbon atoms) comprise a smaller fraction of the oleoresin, yet these compounds are of enormous structural diversity in conifers, exceeding the already impressive diversity of the mono- and diterpenoids (Steele et al., 1998b; Martin et al., 2002). In Norway spruce (*Picea abies*) sesquiterpenoids, along with the monoterpenol linalool, become major constituents of induced volatile emissions following treatment with methyl jasmonate (MeJA; Martin et al., 2003). Whether emitted as volatiles or sequestered in resin ducts, conifer mono-, sesqui-, and diterpenoids occur commonly as complex blends. Their quantity as well as the quality and particular enantiomeric composition contributes to biological activities of terpenoid mixtures in protection against potential pathogens and herbivores.

The basic biosynthetic pathways of terpenoids have been elucidated in several plant systems (Croteau et al.,

[1] This work was supported by the Natural Sciences and Engineering Research Council of Canada (funds to J.B.), by the Canadian Foundation for Innovation and the BC Knowledge and Development Funds (funds to J.B.), and by the Human Frontiers Science Program (funds to J.B.). D.M. is recipient of a Graduate Student Fellowship from the University of British Columbia. J.F. received fellowships from the Bengt Lundqvist Minne Foundation, Sweden, and the Swedish Foundation for International Cooperation in Research and Higher Education.

* Corresponding author; e-mail bohlmann@interchange.ubc.ca; fax 604–822–2114.

[w] The online version of this article contains Web-only data.

Article, publication date, and citation information can be found at www.plantphysiol.org/cgi/doi/10.1104/pp.104.042028.

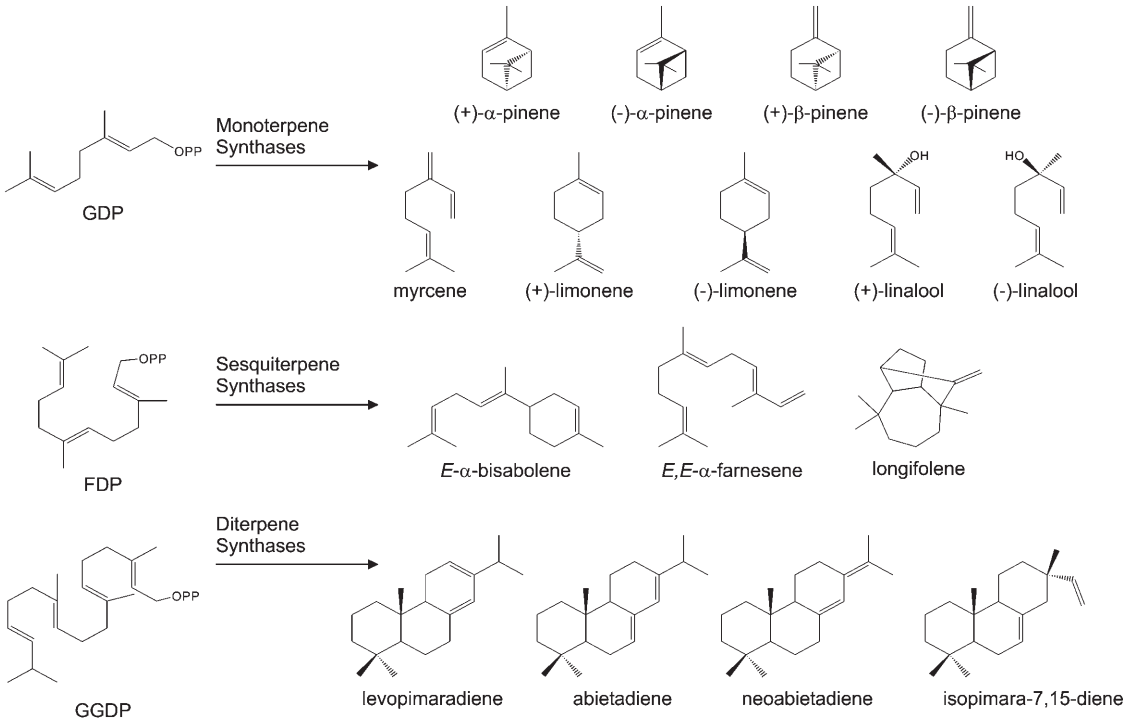

Figure 1. Major Norway spruce TPS products showing monoterpenes, sesquiterpenes, and diterpenes produced by these enzymes.

2000). Geranyl diphosphate (GDP), farnesyl diphosphate (FDP), and geranylgeranyl diphosphate (GGDP) are the immediate precursors to mono-, sesqui-, and diterpenes, respectively (Fig. 1). These allylic prenyldiphosphates are utilized by terpenoid synthases (TPS) of three biochemical classes, monoterpene synthases (mono-TPS), sesquiterpene synthases (sesqui-TPS), and diterpene synthases (di-TPS; Bohlmann et al., 1998b; Davis and Croteau, 2000). TPS form a wide range of structurally diverse cyclic and acyclic mono-, sesqui-, and diterpenes (Fig. 1). TPS function through divalent metal ion dependent generation of enzyme bound carbocation intermediates (Cane, 1999; Wise and Croteau, 1999; Davis and Croteau, 2000). Various mechanisms of rearrangement and quenching of carbocations yield the large array of different terpenoid product profiles of the single-product TPS and multiple-product TPS enzymes (Lesburg et al., 1997; Starks et al., 1997; Cane, 1999; Wise and Croteau, 1999; Davis and Croteau, 2000; Whittington et al., 2002). While many terpenoids exist as pairs of enantiomers (Fig. 1), TPS exhibit preferential catalysis toward one stereoisomer resulting in mostly optically pure product profiles (Davis and Croteau, 2000; Phillips et al., 2003). The exact structural features that determine substrate specificity and the diverse product profiles of TPS enzymes are not known. Thus, biochemical functions of cloned TPS cannot be predicted accurately based on sequence similarity even for closely related enzymes (Bohlmann et al., 1999; Dudareva et al., 2003).

Conifers are a rich source for a large array of TPS that have been characterized primarily from one species, grand fir (*Abies grandis*; Bohlmann and Croteau, 1999). These studies showed that a family of *TPS* genes is central to structural diversity of terpenoid defenses (Stofer Vogel et al., 1996; Bohlmann et al., 1997, 1998a, 1999; Steele et al., 1998a). Recent work in loblolly pine (*Pinus taeda*) provided information about stereospecificity of cloned pinene synthases in this species (Phillips et al., 2003). However, the known grand fir and few loblolly pine *TPS* genes cannot account for all oleoresin terpenoids in conifers. Nothing is known about *TPS* genes of induced volatile emissions in conifers (Martin et al., 2003), and only a few *TPS* genes of herbivore-induced terpene emission are known in other angiosperm plants (Shen et al., 2000; Schnee et al., 2002; Fäldt et al., 2003a; Arimura et al., 2004). Species of spruce (*Picea* spp.), which are among the most important conifers for the forest industry, provide a very interesting system for molecular genetic and functional biochemical characterization of TPS and their roles in terpenoid defenses in gymnosperm trees (Alfaro et al., 2002). In Norway spruce, treatment with MeJA mimics insect attack in the induction of tissue-specific terpenoid responses. These responses include de novo formation of traumatic resin ducts in lieu of tracheids in the developing stem xylem (Franceschi et al., 2002; Martin et al., 2002), induced accumulation of terpenoid resin (Martin et al., 2002), and increased enzyme activities and elevated levels of transcripts of *mono-TPS* and *di-TPS* (Martin

et al., 2002; Fäldt et al., 2003b). *Mono-TPS* transcripts were similarly induced in Sitka spruce (*Picea sitchensis*) by mechanical wounding or feeding by the spruce weevil (*Pissodes strobi*; Byun McKay et al., 2003). MeJA treatment also induced a rhythmic, diurnal release of volatile sesquiterpenoids and oxygenated monoterpenoids, along with methyl salicylate, from needles of Norway spruce (Martin et al., 2003).

Despite increasing knowledge of the anatomical and biochemical processes of induced terpenoid defenses in spruce, detailed molecular and biochemical dissection of these complex defense systems is currently limited by lack of identification and functional characterization of the *TPS* genes involved. Recently, a new *TPS* gene encoding (+)-3-carene synthase was isolated from Norway spruce (Fäldt et al., 2003b) and a (−)-α-pinene synthase gene was identified in Sitka spruce (Byun McKay et al., 2003). However, most spruce *TPS* genes, their particular biochemical functions and their contribution to chemical defense remain to be discovered and functionally characterized. Here we describe cDNA cloning, biochemical characterization, and phylogenetic relatedness of 9 *TPS* genes from Norway spruce, including 4 *mono-TPS*, 3 *sesqui-TPS*, and 2 *di-TPS*. The products of these TPS enzymes are major components of constitutive and MeJA-inducible terpenoid defenses in Norway spruce. We also present maximum likelihood and distance analyses of the gymnosperm TPS-d subfamily where the inclusion of the 10 Norway spruce TPS and 5 additional gymnosperm TPS from the literature made phylogenetic analyses of this subfamily considerably more robust and has provided sufficient evidence that gymnosperm TPS also form phylogenetic clusters based on function and sequence similarities.

RESULTS

cDNA Cloning of a Family of Nine *TPS* Genes from Norway Spruce

We cloned a set of nine *TPS* genes as cDNAs from Norway spruce (Table I). Four of these genes (*PaTPS-Far*, *PaTPS-Lin*, *PaTPS-Bis*, and *PaTPS-Pin*) were isolated by cDNA library filter hybridization as described in Fäldt et al. (2003b). Combination of similarity-based PCR strategies previously developed for isolation of conifer *TPS* genes (Bohlmann et al., 1997), mining of expressed sequence tags, and RACE cloning strategies enabled the isolation of an additional five *TPS* cDNAs (*PaTPS-Lon*, *PaTPS-Lim*, *PaTPS-Iso*, *PaTPS-LAS*, and *PaTPS-Myr*). Based on overall sequence similarity with known *TPS* genes from grand fir and loblolly pine and the Norway spruce (+)-3-carene synthase, the nine TPS cDNAs represented a group of putative mono-TPS, sesqui-TPS, and di-TPS enzymes. Pairwise sequence similarities among predicted amino acids of the nine new *TPS* clones and the Norway spruce (+)-3-carene synthase (Fäldt et al., 2003b) are shown in Table II.

Deduced amino acid sequences (Fig. 2) of the open reading frames of four *TPS* candidate genes, *PaTPS-Lim*, *PaTPS-Myr*, *PaTPS-Lin*, and *PaTPS-Pin*, resembled most closely known conifer mono-TPS (Bohlmann and Croteau, 1999). Their predicted proteins contain in the range of 623 to 634 amino acids and have predicted pI in the range from 5.41 for PaTPS-Myr to 6.51 for PaTPS-Lim. PaTPS-Lin and PaTPS-Pin have very similar pIs at 5.71 and 5.79, respectively. cDNA clones of *PaTPS-Lim*, *PaTPS-Myr*, and *PaTPS-Lin* were full length and each included sequences for a putative N-terminal transit peptide of 62 to 64 amino acids upstream of a conserved RRX_8W motif for import of mature proteins into plastids, characteristic of mono-TPS (Williams et al., 1998; Bohlmann et al., 1998b; Aubourg et al., 2002). *PaTPS-Pin* is slightly truncated, missing a starting Met. The predicted amino acid sequence of PaTPS-Lim is most similar to β-phellandrene synthase (80% I, 89% S) from grand fir (Bohlmann et al., 1999) among the previously identified *TPS* genes. PaTPS-Myr is most similar to myrcene synthase (72% I, 80% S) from grand fir (Bohlmann et al., 1997). The third mono-TPS of this set, PaTPS-Pin, is most similar to (−)-pinene synthase (90% I, 91% S) from Sitka spruce (Byun McKay et al., 2003). The amino acid sequence of PaTPS-Lin is closest to the previously characterized terpinolene synthase (69% I, 81% S) from grand fir (Bohlmann et al., 1999).

Table I. *Gene name, accession numbers, and functional annotation of Norway spruce TPS*

Gene	Clone	Accession No.	TPS Class	Functional Annotation[a]
PaTPS-Car	PaJF67	AF461459	Mono-TPS	(+)-3-Carene synthase
PaTPS-Lim	PaDM743	AY473624	Mono-TPS	(−)-Limonene synthase
PaTPS-Myr	PaJB16	AY473626	Mono-TPS	Myrcene synthase
PaTPS-Lin	PaJF39	AY473623	Mono-TPS	(−)-Linalool synthase
PaTPS-Pin	PaJF104	AY473622	Mono-TPS	(−)-α/β-Pinene synthase
PaTPS-Far	PaJF71	AY473627	Sesqui-TPS	*E,E*-α-Farnesene synthase
PaTPS-Bis	PaDM03	AY473619	Sesqui-TPS	*E*-α-Bisabolene synthase
PaTPS-Lon	PaDM486	AY473625	Sesqui-TPS	Longifolene synthase
PaTPS-LAS	PaDM2420	AY473621	Di-TPS	Levopimaradiene/abietadiene s.
PaTPS-Iso	PaDM2425	AY473620	Di-TPS	Isopimara-7,15-diene synthase

[a]Functional annotation is based on the main terpenoid product of recombinant enzymes expressed in *E. coli*. Several PaTPS form multiple products which are described in the text and shown in the corresponding figures with GC-MS profiles.

Table II. *Sequence relatedness of Norway spruce TPS*

Results from pairwise amino acid sequence comparisons are shown as percent identity (upper line) and percent similarity (lower line) for mono-TPS (black, italics), sesqui-TPS (black, bold), and di-TPS (underline) from Norway spruce.

	a	b	c	d	e	f	g	h	i	j
(−)-Limonene synthase (a)	*100*									
	–									
Myrcene synthase (b)	*69*	*100*								
	80	–								
(−)-α/β-Pinene synthase (c)	*68*	*65*	*100*							
	83	*80*	–							
(−)-Linalool synthase (d)	*66*	*61*	*68*	*100*						
	82	*75*	*81*	–						
(+)-3-Carene synthase (e)	*66*	*61*	*65*	*68*	*100*					
	81	*75*	*80*	*82*	–					
***E,E*-α-Farnesene synthase (f)**	***62***	***67***	***62***	***64***	***60***	**100**				
	75	***79***	***76***	***76***	***72***	–				
Longifolene synthase (g)	***38***	***43***	***40***	***39***	***39***	**40**	**100**			
	57	***63***	***58***	***59***	***58***	**61**	–			
***E*-α-Bisabolene synthase (h)**	***29***	***29***	***29***	***28***	***28***	**29**	**30**	**100**		
	42	***43***	***42***	***42***	***42***	**43**	**44**	–		
Levopimaradiene/abietadiene synthase (i)	*25*	*25*	*25*	*24*	*24*	24	27	42	100	
	38	*38*	*37*	*38*	*37*	37	37	61	–	
Isopimara-7,15-diene synthase (j)	*24*	*25*	*24*	*24*	*24*	23	26	41	90	100
	37	37	36	*37*	*36*	36	37	60	94	–

Three Norway spruce TPS clones, *PaTPS-Far*, *PaTPS-Lon*, and *PaTPS-Bis*, resemble conifer sesqui-TPS, although they are quite different among each other in sequence similarity and length and represent the three different structural types of known conifer sesqui-TPS enzymes (Bohlmann et al., 1998a; Steele et al., 1998a; Phillips et al., 2003). All three predicted amino acid sequences lack long N-terminal sequences upstream of the RRX_8W motif (Fig. 2)—consistent with cytsolic localization of plant sequi-TPS—and each had a predicted pI near 5.3. PaTPS-Far and PaTPS-Lon have predicted lengths of 580 and 578 amino acids. PaTPS-Bis is longer at 807 amino acids due to an additional sequence motif of approximately 200 amino acids, similar to the ancestral di-TPS motif (Bohlmann et al., 1998a, 1998b). Of these three putative sesqui-TPS, the amino acid sequence of PaTPS-Far has highest similarity with the loblolly pine α-farnesene synthase (81% I, 90% S; Phillips et al., 2003), but surprisingly also has very high sequence similarity with the spruce mono-TPS described above (Table II). PaTPS-Bis is most similar to grand fir *E*-α-bisabolene synthase (86% I, 91% S; Bohlmann et al., 1998a). The third putative sesqui-TPS, PaTPS-Lon, is most similar with δ-selinene synthase from grand fir (62% I, 82% S; Steele et al., 1998a).

Two TPS cDNAs, *PaTPS-LAS* and *PaTPS-Iso*, were very similar to the known conifer di-TPS, abietadiene synthase from grand fir (Stofer Vogel et al., 1996) and taxadiene synthase from *Taxus brevifolia* (Wildung and Croteau, 1996). Each contained the ancestral conifer di-TPS motif of 200 to 215 amino acids of unknown function (Bohlmann et al., 1998b), which makes these proteins with approximately 860 amino acids substantially longer than most mono-TPS or sesqui-TPS (Fig. 2). Both genes encoded for a putative transit peptide upstream of the amino acid sequence KREFPPGFW (Fig. 2). This motif is very similar in all known conifer di-TPS and is positionally and functionally conserved with the RRX_8W motif in mono-TPS (Bohlmann et al., 1998b).

Functional Expression and Biochemical Characterization of Cloned TPS Enzymes

Recombinant TPS enzymes were functionally characterized on the basis of substrate specificity, product profile and chirality of terpenoids formed. To determine the function of each of these *TPS* genes, all cDNAs were subcloned into expression vectors, expressed in *Escherichia coli*, and cell-free extracts of all recombinant enzymes were tested for activity with the substrates GDP, FDP, and GGDP (Bohlmann et al., 1997; Peters et al., 2000; Fäldt et al., 2003b). In some cases, it was necessary to try several different vectors to optimize expression of active protein. We used a set of four pET-derived vectors for expression of spruce *TPS* cDNAs including the pSBET vector that codes for additional Arg tRNAs and was previously shown to enhance expression of active TPS enzymes in *E. coli* (Bohlmann et al., 1999), and the pET100/D, pET101/D, and pQE50 vectors, which were previously shown to work effectively with *TPS* expression (Miller et al., 2001; Fäldt et al., 2003b). Terpenoid products formed by the recombinant TPS enzymes (Table III) were analyzed using gas chromatography (GC) and mass spectrometry (MS) and/or GC-flame ionization detection (GC-FID). Since many of these spruce TPS

have complex product profiles, approximate relative amounts of individual compounds were calculated from GC-MS analyses where the fragmentation patterns assisted in product identification. Terpene identification was further validated by comparing retention times and mass spectra with those of authentic standards and library matches.

Functional Identification of Four *Mono-TPS* Genes

Four of the nine *TPS* cDNA clones, *PaTPS-Lim*, *PaTPS-Myr*, *PaTPS-Lin*, and *PaTPS-Pin*, were confirmed after expression and functional characterization to encode mono-TPS enzymes (Table III; Fig. 3; Supplemental Figs. 1–3, which can be viewed at www.plantphysiol.org). The product profile of each mono-TPS was initially identified by GC-MS using an HP-5 or a DB-Wax column. Further analysis was performed to identify monoterpene enantiomers (Fig. 3; Supplemental Figs. 1–3). Of the four mono-TPS, three enzymes encoded by *PaTPS-Lim*, *PaTPS-Myr*, and *PaTPS-Pin* form monoterpene hydrocarbons. PaTPS-Lim formed predominantly (−)-limonene (87.8%) along with four other minor products when incubated with GDP and was designated (−)-limonene synthase (Table III; Supplemental Fig. 1). PaTPS-Myr catalyzed the formation of a single compound, myrcene, when incubated with GDP thereby designating this TPS as a myrcene synthase (Table III; Supplemental Fig. 2). Chiral analysis of products formed by PaTPS-Pin with GDP as substrate revealed (−)-β-pinene (57.1%), (−)-α-pinene (27.4%), and (−)-β-phellandrene (11.0%) as the major monoterpenes produced along with five other minor constituents (Table III; Supplemental Fig. 3). This TPS has been named (−)-α/β-pinene synthase. Assays with GDP and product analysis of PaTPS-Lin revealed a profile of two monoterpene alcohols, (−)-linalool (96.6%) and (+)-linalool (1.6%; Fig. 3), and an additional nine monoterpene hydrocarbons at very low abundance including *E*-β-ocimene, myrcene, α-terpinolene, and 3-carene (Table III). This is the first linalool synthase isolated from a gymnosperm. The near optical purity of (−)-linalool synthase indicates that active site residues exert considerable control over the quenching of the carbocation by the addition of water.

Functional Identification of Three *Sesqui-TPS* Genes

Three cDNA clones, *PaTPS-Far*, *PaTPS-Bis*, *PaTPS-Lon*, were functionally identified as sesqui-TPS. After assaying PaTPS-Far with each of the three prenyl diphosphate substrates, GDP, FDP, and GGDP, activity was seen only with FDP. This enzyme produces *E,E*-α-farnesene as a single product (Table III; Supplemental Fig. 4), confirming its identity as a sesqui-TPS. We have designated this TPS as *E,E*-α-farnesene synthase. Expression of PaTPS-Bis and enzyme assays revealed *E*-α-bisabolene as the single product of this enzyme identifying it as *E*-α-bisabolene synthase (Table III; Supplemental Fig. 5). The third sesqui-TPS, PaTPS-Lon, produced mainly longifolene (60.6%; Fig. 4, Table III) when assayed with FDP. This sesqui-TPS also produced multiple minor products including α-longipinene, α-longicyclene, *E*-β-farnesene, the sesquiterpene alcohol, longiborneol, cyclosativene, β-longipinene, and 12 other sesquiterpenes (each under 1.4% of total product). This enzyme was designated longifolene synthase for its major sesquiterpene product.

Functional Identification of Two *Di-TPS* Genes

The first of two Norway spruce *di-TPS* candidate cDNAs, *PaTPS-LAS*, encodes for a predicted protein of 859 amino acids (Fig. 2) and a pI of 5.52. Several truncations were made of this cDNA to determine if expression was improved with the removal of transit peptide (Peters et al., 2000); however the full-length gene gave the highest di-TPS activity. When incubated with GGDP, PaTPS-LAS formed the four diterpenes, levopimaradiene (36.5%, peak 1), abietadiene (32.1%, peak 2), neoabietadiene (23.2%, peak 3) and palustradiene (8.1%, peak 4; Fig. 5, A–D; Table III). We used grand fir abietadiene synthase (Stofer Vogel et al., 1996; Peters et al., 2000) to compare product profiles of these two similar enzymes and found that product profiles of the two enzymes were nearly identical to each other and to published results with grand fir abietadiene synthase (Peters et al., 2000). We have designated this di-TPS levopimaradiene/abietadiene synthase (PaTPS-LAS) after the major products it forms. The second di-TPS, isopimara-7,15-diene synthase (PaTPS-Iso), is slightly longer than PaTPS-LAS at a predicted length of 867 amino acids (Fig. 2). The predicted protein pI is 5.83. Heterologous expression of full-length PaTPS-Iso and enzyme assays with GGDP yielded a single product, isopimara-7,15-diene (Fig. 5, E–F; Table III). This di-TPS was hence designated PaTPS-Iso. To our knowledge, an PaTPS-Iso has not been previously described for any species.

The proposed reaction mechanisms for PaTPS-LAS and PaTPS-Iso proceed from *E,E,E*-GGDP to the bicyclic (+)-copalyl diphosphate intermediate (Fig. 6). Studies in grand fir abietadiene synthase have demonstrated that this intermediate then freely diffuses to a second active site of the bifunctional enzyme where the sandaracopimarenyl intermediate is produced (Peters et al., 2001, 2003). Following the generation of sanadaracopimarenyl carbocation, the mechanisms of PaTPS-LAS and PaTPS-Iso diverge. To form the four products of LAS, a 1,2-methyl migration occurs to yield the abietane structure followed by deprotonation to give the different double bond configurations of levopimaradiene, abietadiene, neoabietadiene, and palustradiene. The proposed reaction scheme of PaTPS-Iso involves simple deprotonation of the sandaracopimarenyl carbocation (Fig. 6).

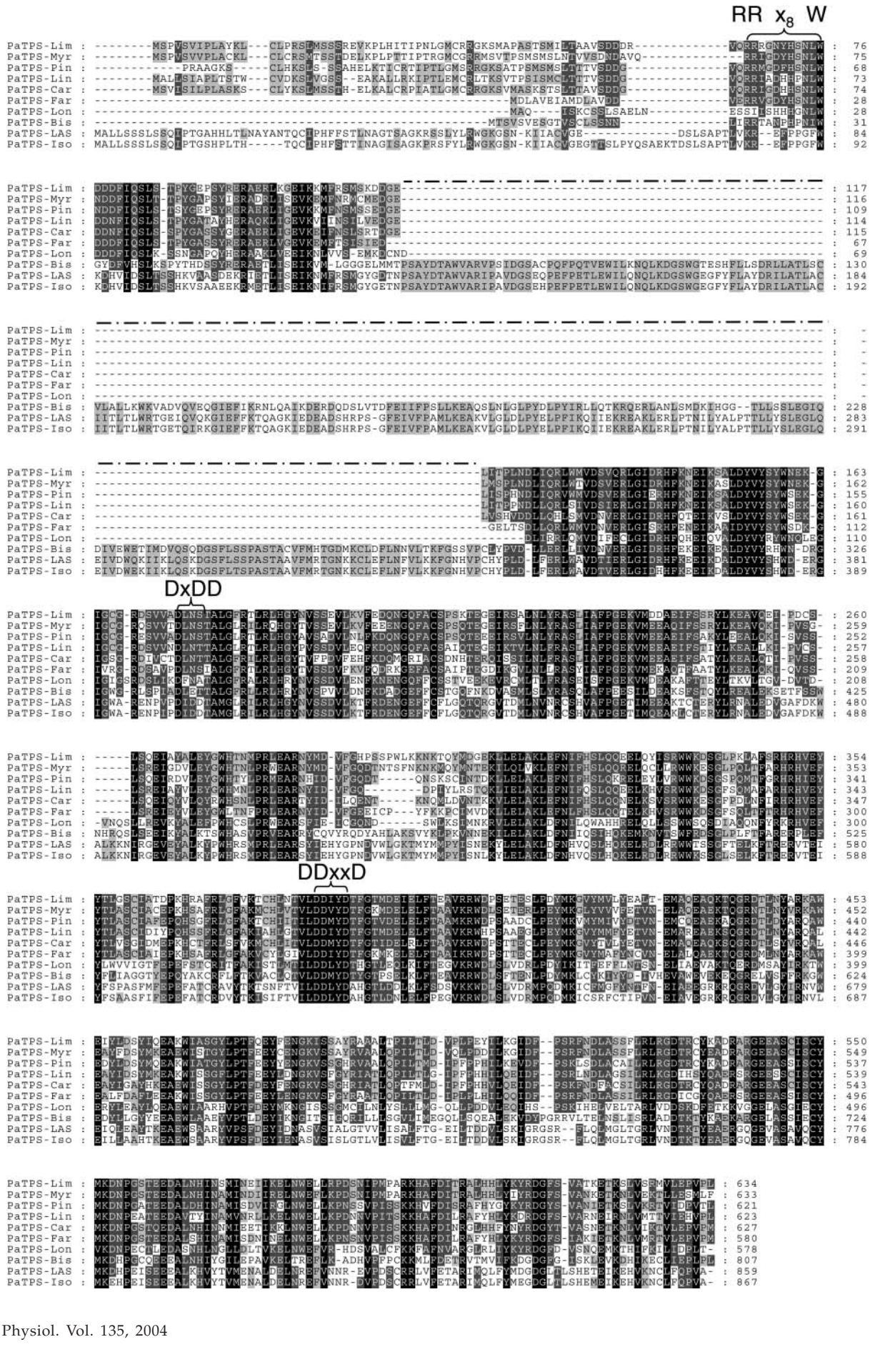
RR x8 W
DxDD
DDxxD

Table III. *Norway spruce mono-, sesqui-, and diterpene synthases and the products that each produces*

TPS	Products	Percent Total
	Monoterpene Synthases	
PaTPS-Lim	(−)-Limonene	87.8
	Myrcene	5.2
	(−)-α-Pinene	4.4
	(+)-Limonene	2.1
	(−)-β-Pinene	0.5
PaTPS-Myr	Myrcene	100
PaTPS-Pin	(−)-β-Pinene	57.1
	(−)-α-Pinene	27.4
	(−)-β-Phellandrene	11
	(+)-β-Pinene	0.6
	(−)-α-Pinene	0.7
	Myrcene	2.1
	(−)-Limonene	0.4
	(+)-Limonene	0.7
PaTPS-Lin	(−)-Linalool	96.6
	(+)-Linalool	1.6
	E-β-Ocimene	1.0
	Myrcene	0.2
	α-Terpinolene	0.1
	3-Carene	0.2
	Other monoterpenes	Each <0.07
	Sesquiterpene Synthases	
PaTPS-Far	*E,E*-α-Farnesene	100
PaTPS-Bis	*E*-α-Bisabolene	100
PaTPS-Lon	Longifolene	60.6
	α-Longipinene	14.6
	Longicyclene	5.9
	E-β-Farnesene	3.4
	Longiborneol	2.1
	Cyclosativene	1.3
	β-Longipinene	1.4
	Other sesquiterpenes	Each <1.4
	Diterpene Synthases	
PaTPS-LAS	Levopimaradiene	36.5
	Abietadiene	32.1
	Neoabietadiene	23.2
	Palustradiene	8.1
PaTPS-Iso	Isopimara-7,15-diene	100

Phylogeny of the Gymnosperm TPS-d Subfamily

The monophyletic plant TPS family has been delineated into separate subfamilies (TPS-a through TPS-g) based on sequence relatedness as well as functional assessment (Bohlmann et al., 1998b). The extant TPS of plant secondary metabolism apparently arose from an ancestral *TPS* gene that was likely most similar in gene structure and sequence to the known gymnosperm di-TPS (Trapp and Croteau, 2001b; Bohlmann et al., 1998b). Such an ancestral *di-TPS* gene may have been involved in GGDP cyclization of gibberellic acid metabolism. While all extant *TPS* genes characterized to date seem to share a common origin, functional specialization of many TPS occurred after the separation of angiosperms and gymnosperms (Bohlmann et al., 1998b; Trapp and Croteau, 2001b).

All known gymnosperm TPS cluster in the TPS-d subfamily (Fig. 7 and Fig. 8) which contains mono-, sesqui-, and di-TPS. Prior to the cloning of Norway spruce TPS, the gymnosperm TPS-d subfamily consisted mostly of TPS from grand fir (Bohlmann et al., 1998b), a member of the pine family (Pinaceae). Sufficient information on other conifer species was lacking thereby preventing a thorough comparison of TPS from different taxa of the pine family and an understanding of TPS evolution within the gymnosperm TPS-d subfamily, and how this might compare with TPS evolution in angiosperms. For instance, it was not known whether distinct branching patterns would further dissect the TPS-d subfamily based on function or whether the functional speciation of TPS preceded the speciation of the Pinaceae. However, with the inclusion of 10 new Norway spruce TPS as well as some gymnosperm TPS recently available from the literature, we can now reconstruct and further understand the phylogeny of the TPS-d subfamily. To this goal, we performed a maximum likelihood analysis of conifer TPS using PHYML (Guindon and Gascuel, 2003) and the JTT substitution matrix (Jones et al., 1992). The alpha shape parameter as well as the invariable site parameter was estimated prior to analysis of 100 bootstrap replicates. Phylogenetic trees were constructed using the modified neighbor joining program BIONJ (Gascuel, 1997). Bootstrap values are also included for distance analysis using PROTDIST (Felsenstein, 1993) and the Dayhoff PAM 001 matrix (Dayhoff, 1979).

The inclusion of 17 additional conifer TPS, including 10 from Norway spruce, 3 mono-TPS and 1 sesqui-TPS from loblolly pine, 2 mono-TPS from Sitka spruce, and 1 di-TPS from *Ginkgo biloba* (Supplemental Table I), fleshed out the resulting tree such that we can now discern 3 discrete groups within the gymnosperm TPS-d subfamily (Fig. 7). As the identity of the TPS-d subfamily should be maintained, we propose the labeling for these groups to be TPS-d1, TPS-d2, and TPS-d3, based on sequence relatedness and functional assessment. Each group is dominated by TPS of 1 of the 3 biochemical classes, mono-TPS (TPS-d1), sesqui-TPS (TPS-d2), and di-TPS (TPS-d3). This grouping is analogous to the overall delineation of angiosperm

Figure 2. Amino acid alignments of Norway spruce mono-TPS, sesqui-TPS, and di-TPS generated by ClustalX and GeneDoc. The RRX_8W, common to nearly all mono-TPS and positionally preserved in some sesqui-TPS and di-TPS. Asp rich motifs, DxDD for di-TPS, and DDxxD for mono-TPS, sesqui-TPS, and di-TPS—necessary for the binding of cationic cofactors—are shown. The ancestral conifer di-TPS motif of 200 to 215 amino acids is delineated by a dashed and dotted line. Conserved similarity shading is based on 100% (black), 60% (dark gray), and 30% (light gray).

TPS where each subfamily is composed of TPS with similar sequence structures and, in most cases, angiosperm TPS subfamilies contain members of the same biochemical class (Bohlmann et al., 1998b).

Considering the extremely long evolutionary divergence of extant gymnosperms, it is a striking feature of this analysis that most of the known gymnosperm TPS still segregate within the tree by common overall function. For example, all conifer mono-TPS fall into the TPS-d1 cluster and all di-TPS cluster in the TPS-d3 group (Fig. 7). In contrast to mono-TPS, fewer sesqui-TPS have been identified in gymnosperms. While conifer sesqui-TPS form the TPS-d2 group, other sesqui-TPS cluster with the mono-TPS or di-TPS suggesting multiple origins of sesqui-TPS functions (see below).

The fact that all di-TPS cluster closely together, including the levopimaradiene synthase from G. *biloba*, taxadiene synthase from Taxus, and several conifer di-TPS, infers that the initial duplication event and inception of di-TPS of specialized gymnosperm secondary metabolism occurred before the separation of Coniferales and Ginkgoales probably more than 250 to 290 million years ago. Another important point gathered from this analysis includes the presence of intraspecies gene duplications followed by divergent functional evolution. This is illustrated by the paralogs PaTPS-Iso and PaTPS-LAS from Norway spruce (Fig.

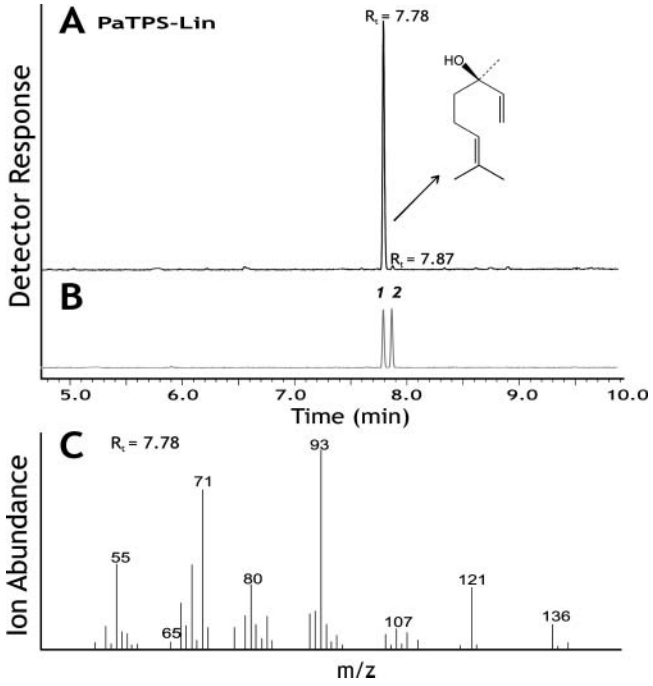

Figure 3. GC-MS chiral analysis of products formed by PaTPS-Lin, (−)-linalool synthase. Total ion chromatogram of assay products (A) showing a major peak, R_t = 7.78 along with (B) chiral terpene standards showing (−)-linalool, R_t = 7.79 (1), and (+)-linaool, R_t = 7.87 (2). Mass spectra of major peak (C) identifies this peak as (−)-linalool.

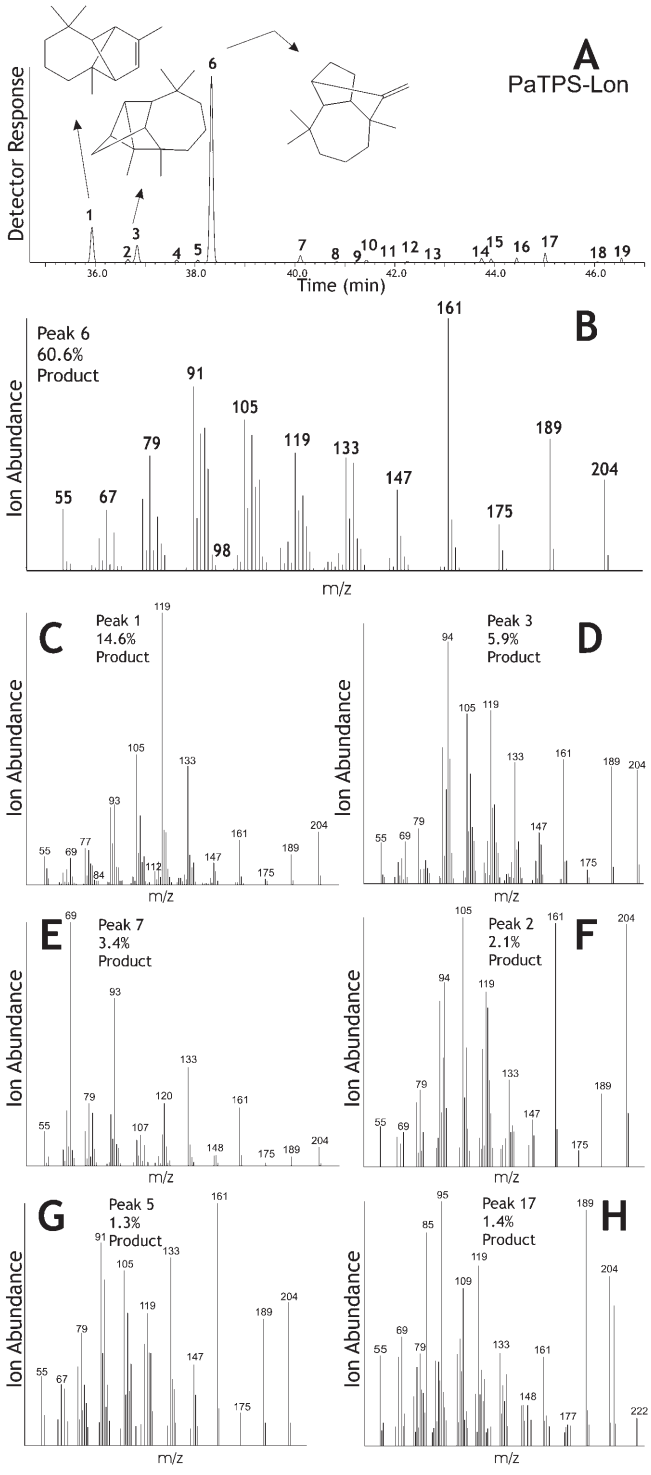

Figure 4. GC-MS analysis of the multiple-product forming PaTPS-Lon, longifolene synthase. Total ion chromatogram of the assay showing 19 sesquiterpene products (A). Mass spectra of the major peak (B) and six additional products according to decreasing abundance including α-longicyclene (C), α-longipinene (D), *E*-β-farnesene (E), cyclostativene (F), β-longipinene (G), and longiborneol (H).

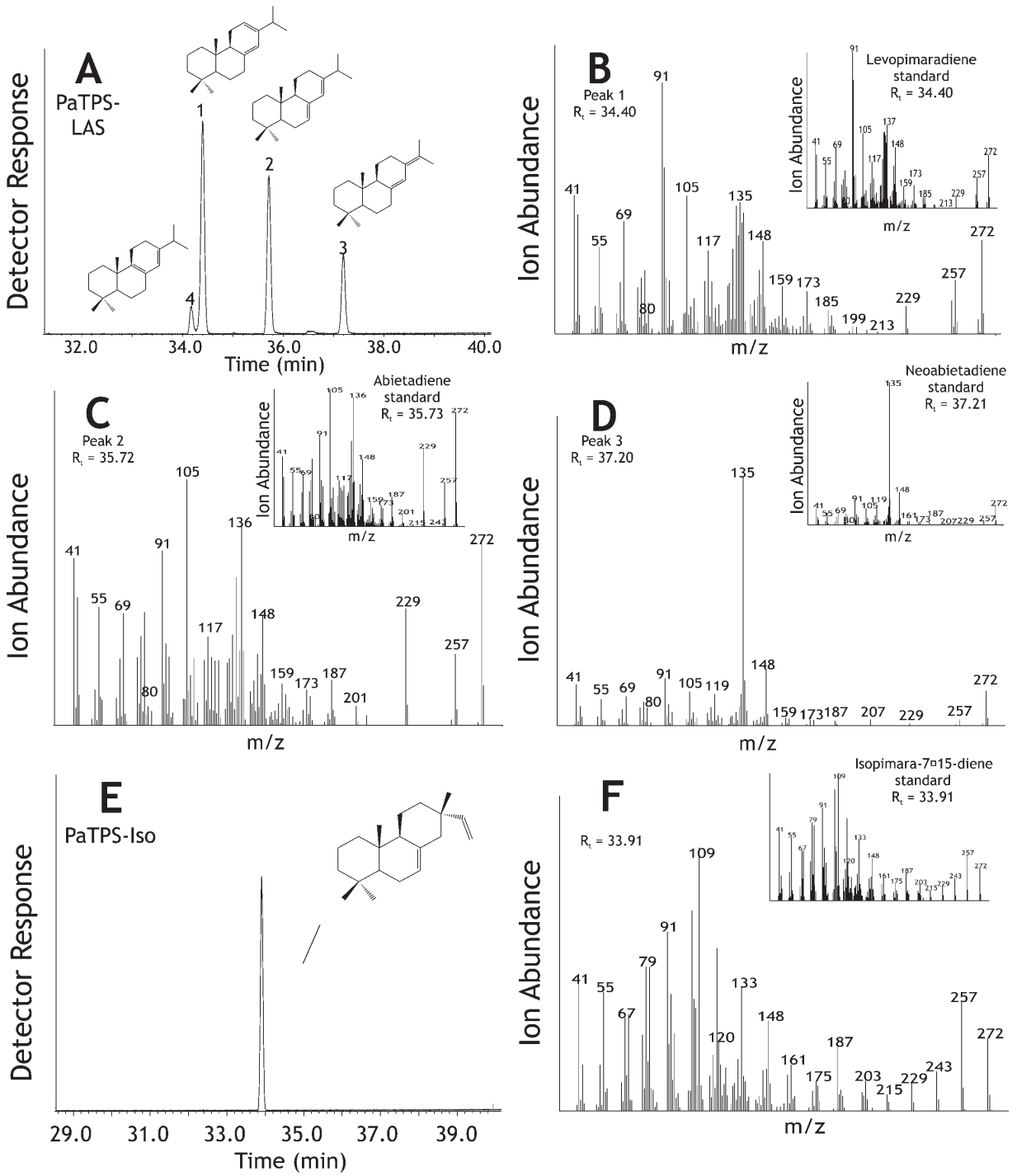

Figure 5. GC-MS analysis of products formed by PaTPS-LAS and PaTPS-Iso. Total ion chromatogram of multiple assay products of PaTPS-LAS (A). Mass spectra are shown of peak one with insert of levopimaradiene standard (B), peak two with insert of authentic abietadiene (C) and of peak three with insert of neoabietadiene standard (D). Total ion chromatogram of the single product produced by PaTPS-Iso (E). Mass spectra of this product are shown with an insert of authentic isopimara-7,15-diene (F).

7). While these two proteins are 94% similar and 90% identical, they produce very different, nonoverlapping diterpenoid profiles (Fig. 6; Table III), which makes this pair of proteins an interesting target for future, biochemical structure-function analysis.

Another distinct result from this phylogeny of gymnosperm TPS is that a large suite of TPS with specific biochemical functions and substrate preferences apparently arose before the speciation of members of the Pinaceae and remained surprisingly stable since

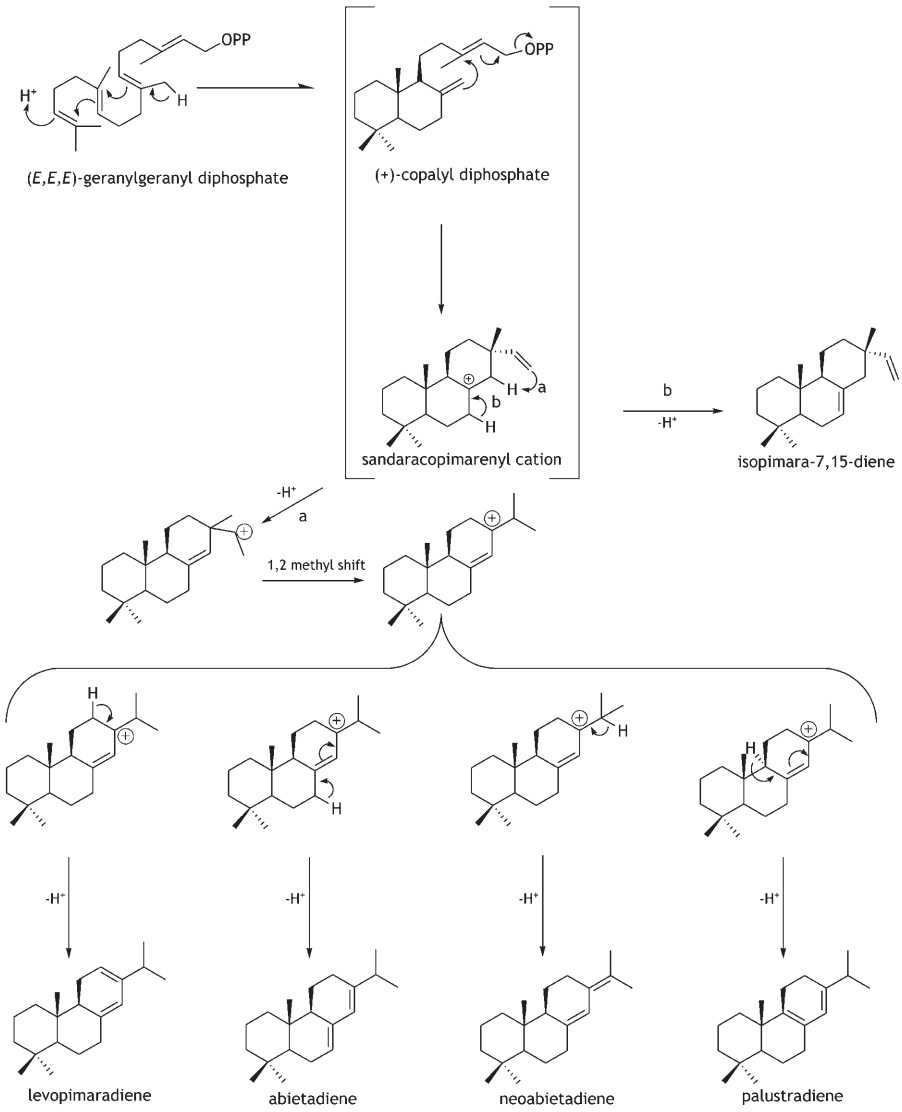

Figure 6. Proposed reaction mechanisms for diterpene synthases PaTPS-LAS and PaTPS-Iso. The substrate GGDP undergoes initial cyclization reaction yielding (+)-copalyl diphosphate. In a second cyclization reaction, the intermediate sandaracopimarenyl carbocation is formed. The pathway to the left (a) reflects the reaction mechanism of PaTPS-LAS whereby a 1,2 methyl migration results in the abietane skeleton. Final deprotonations from this skeleton yield the four detected products, levopimaradiene, abietadiene, neoabietadiene, and palustradiene, each with different double bond configurations. The pathway toward the right (b) indicates the reaction mechanism of PaTPS-Iso whereby the sandaracodimarenyl carbocation is deprotonated to yield isopimara-7,15-diene.

then. Examples of this include myrcene synthases, *E*-α-bisabolene synthases, and the pair of PaTPS-LAS and abietadiene synthase, each represented by orthologous gene pairs of the exact same biochemical function in Norway spruce and grand fir (Fig. 7; Supplemental Table II), and all supported by high bootstrap values. Similarly, loblolly pine and Norway spruce α-farnesene synthases also seem to be a very old and functionally conserved orthologous gene pair, since pines and spruce represent taxa of the Pinaceae that separated around 140 million years ago (X.Q. Wang et al., 2000). Yet other TPS appear to have independently evolved further since speciation of members of the pine family such that the product profiles of these enzymes have more or less diversified though these genes are certainly homologous. The (−)-α-pinene synthases from Norway spruce, Sitka spruce, loblolly pine, and grand fir are good examples of how some TPS have evolved to produce additional new products or different product ratios (Fig. 7; Supplemental Table II). The (−)-pinene synthases of Norway spruce and grand fir both produce a greater proportion of (−)-β-pinene than (−)-α-pinene (Bohlmann et al., 1997), whereas in the closely related (−)-α-pinene synthase of Sitka spruce, (−)-α-pinene is the major product (Byun McKay et al., 2003) and in loblolly pine(−)-α-pinene is the only product (Phillips et al., 1999; Supplemental Table II). Notably, each of these still forms (−)-α-pinene as a major product and they all group closely together within the TPS-d1 cluster (Fig. 7). A similar situation exists for the (−)-limonene synthases from Norway spruce (this study), grand fir (Bohlmann et al., 1997), and Sitka spruce (K.-A. Godard, S. A. Byun McKay, L.A. Plant, and J. Bohlmann, unpublished data). These TPS homologs are all examples of how TPS can evolve to produce new functional variations on overall similar product profiles (Supplemental Table II). The phylogeny of these homologs is also supported by high bootstrap values (Fig. 7). For some TPS such as α-terpineol synthase, (+)-3-carene synthase, and

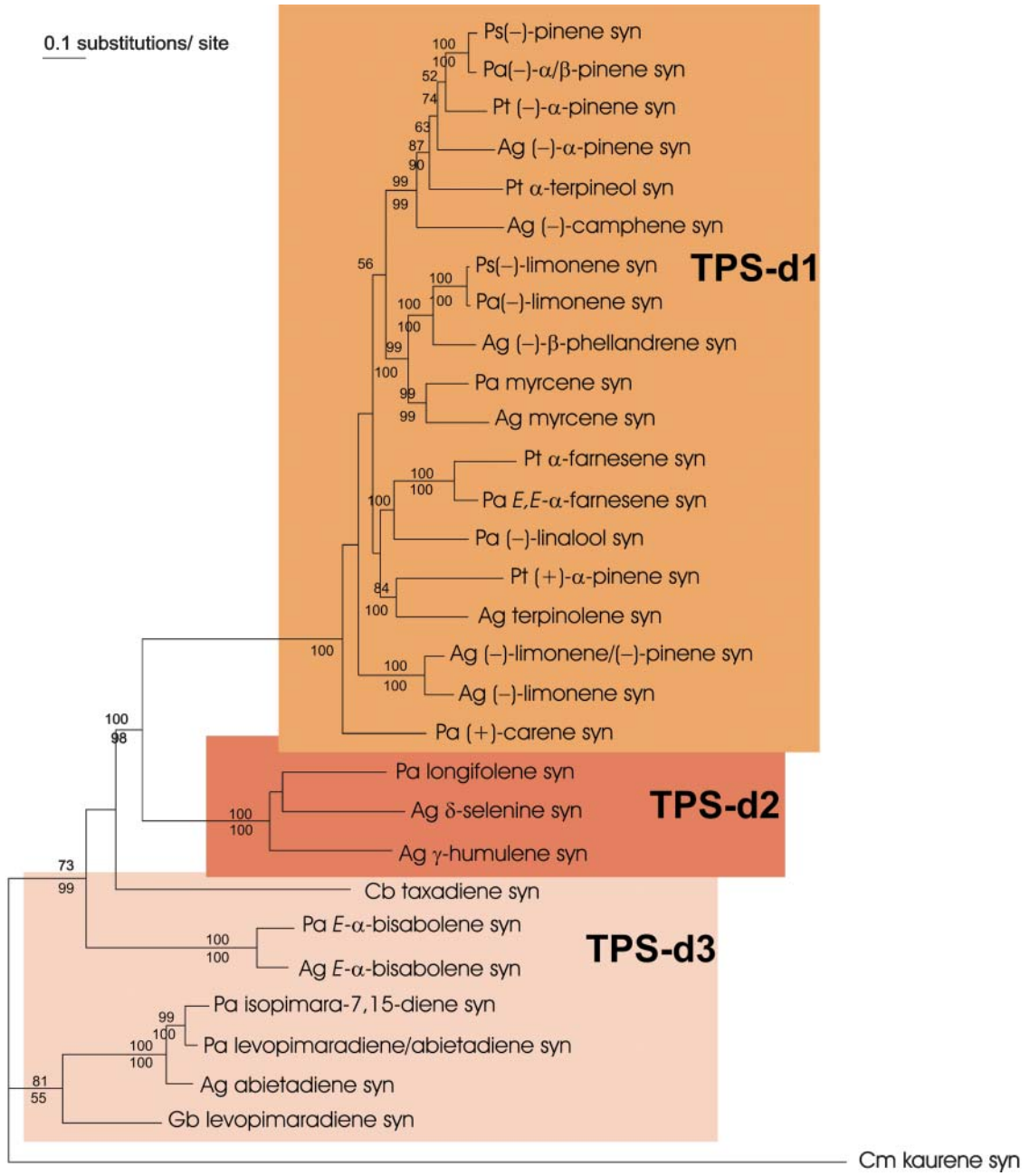

Figure 7. Phylogenetic tree of gymnosperm TPS amino acid sequences (Supplemental Table I) showing the TPS-d subfamily branching into three distinct groups of TPS involved in secondary metabolism. Branching patterns seen in this analysis are cause for the further separation into TPS-d1 (primarily mono-TPS), TPS-d2 (sesqui-TPS), and TPS-d3 (primarily di-TPS). Kaurene synthase from *Cucurbita maxima* shown as outgroup in this analysis. Bootstrap values over 50% for maximum likelihood (upper) and distance (lower) analyses are shown at nodes. Maximum likelihood values represent percentages of 100 γ-corrected (log L = −52520.29) replicates analyzed using Phyml.

(−)-camphene synthase additional genes need to be identified before we begin to understand their evolution within the gymnosperms better. Finally, some TPS functions [e.g. (−)-limonene synthase] seem to have evolved more than once within the TPS-d1 subfamily.

Despite the much refined analysis of the gymnosperm TPS-d family presented here, the phylogenetic position of a gymnosperm di-TPS of gibberellic acid phytohormone metabolism relative to di-TPS of secondary metabolism remains enigmatic until a gymnosperm (−)-copalyl diphosphate synthase or kaurene

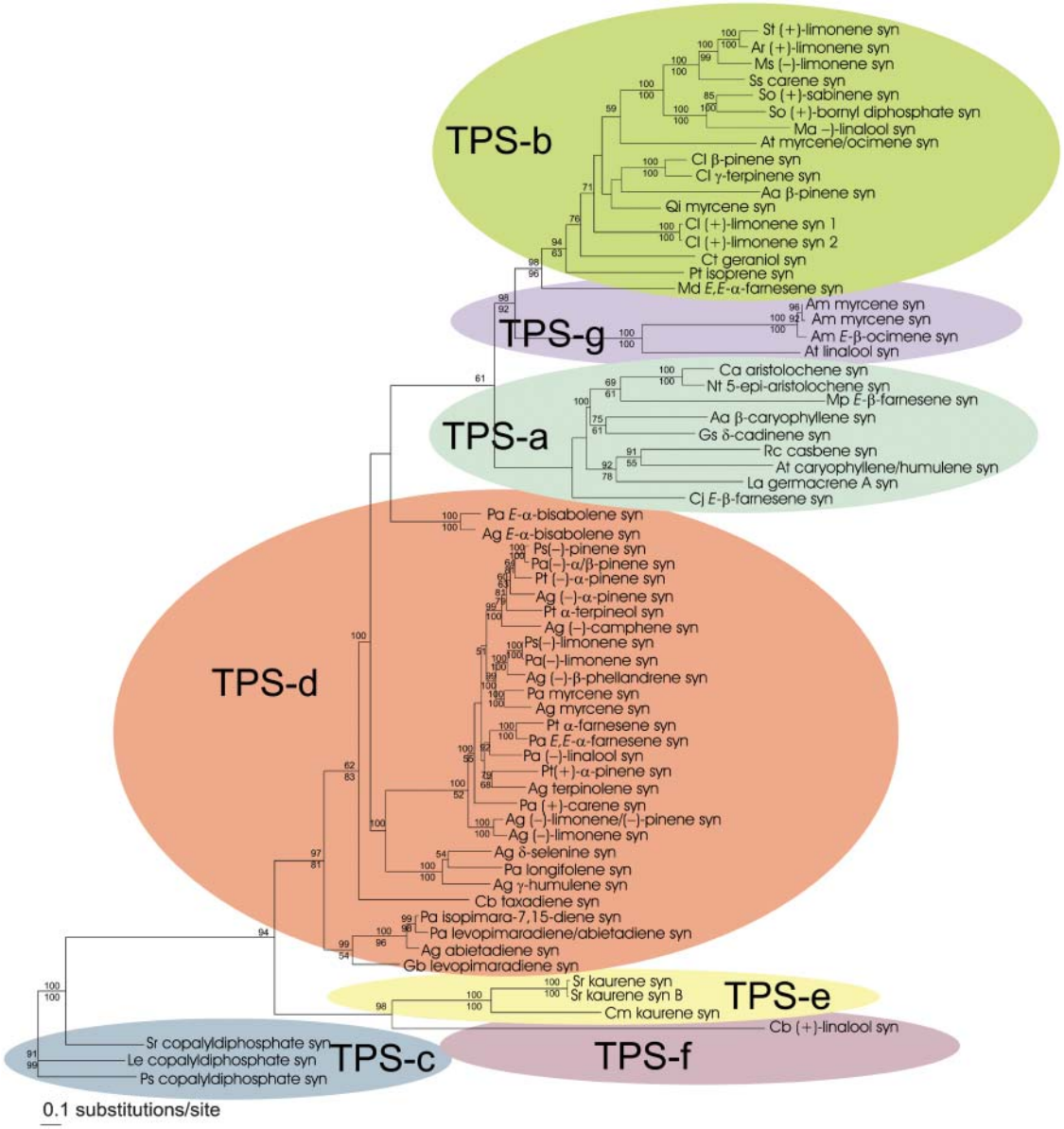

Figure 8. Phylogenetic tree illustrating the relationship of TPS involved in primary and secondary metabolism from angiosperms and gymnosperms (Supplemental Table I). Amino acids sequences of 67 TPS were analyzed by maximum likelihood using Phyml. Bootstrap values over 50% present at nodes for maximum likelihood (upper) and distance (lower) analyses. Maximum likelihood values represent percentages of 100 gamma-corrected (log L = −20452.16) replicates.

synthase is cloned. It is indeed surprising that these gene functions have not yet been discovered in a gymnosperm despite considerable expressed sequence tag resources from pines and spruce and efforts of targeted screening for conifer di-TPS.

Comparison of Phylogeny of TPS from Angiosperms and Gymnosperms

After visualizing the placement of the newly identified Norway spruce TPS within the TPS-d subfamily, we wanted to further understand how these TPS related to the larger family comprised of both angiosperm and gymnosperm TPS. It was previously established that extant TPS functions evolved independently in gymnosperms and angiosperms (Bohlmann et al., 1998b). The numerous known angiosperm TPS of both primary and secondary metabolism are located within six subfamilies (Bohlmann et al., 1998b; Dudareva et al., 2003) that are distinct from the gymnosperm TPS-d cluster. Angiosperm TPS subfamilies are delineated into sesqui-TPS (TPS-a), mono-TPS (TPS-b), copalyl diphosphate synthases (TPS-c), and ent-kaurene synthases (TPS-e) of gibberellic acid

formation and secondary metabolism, the TPS-f subfamily with *Clarkia breweri* linalool synthase, and the newly identified TPS-g cluster, which consists of a group of mono-TPS that produce acyclic compounds. We chose representative TPS from each of the angiosperm TPS subfamilies (Supplemental Table I) and the TPS-d subfamily and analyzed them as described above. Individual TPS subfamilies (TPS-a through TPS-g) are clearly demarcated within this analysis (Fig. 8). In agreement with the analysis of genomic TPS sequences (Trapp and Croteau, 2001b), the di-TPS of the gymnosperm TPS-d subfamily cluster near the angiosperm di-TPS of gibberellic acid metabolism and the *C. breweri* linalool synthase. The close proximity of these sequences is largely influenced by the presence of the ancestral approximately 200-amino acid motif, an attribute that is shared among all these di-TPS sequences, *C. breweri* linalool synthase, and several sesqui-TPS of the TPS-d subfamily, but apparently has been lost in the TPS-a, TPS-b, and TPS-g subfamilies as well as in all conifer mono-TPS and most conifer sesqui-TPS known to date (Bohlmann et al., 1998b; Trapp and Croteau, 2001b).

An intriguing new discovery from comparative analysis of gymnosperm and angiosperm involves the presence of a cluster of unusual TPS that form either acyclic sesquiterpenes, oxygenated monoterpenoids, or the hemiterpene isoprene within the TPS-d1 and the TPS-b subfamilies that are otherwise dominated by regular mono-TPS (Fig. 8). Surprisingly, the two angiosperm and gymnosperm mono-TPS subfamilies both contain farnesene synthases. The independent characterization of ortholog farnesene synthases from Norway spruce (this study) and from loblolly pine (Phillips et al., 2003) provides good evidence that these unexpected sesquiterpene synthase activities are not the result of cloning or other artifacts. However, this finding indicates that a relatively simple sesqui-TPS enzyme mechanism for the formation of acyclic farnesene from FDP likely arose from mono-TPS precursor enzymes in both angiosperms and gymnosperms. At least two necessary events were required for this to take place as exons coding for transit peptides must have been eliminated, and modifications of the original mono-TPS active site enabling the acceptance of FDP over GDP must have occurred. Norway spruce farnesene synthase was not active with GDP, suggesting that this is not the result of simple extension of substrate specificity but involves the more difficult to explain loss of activity with the smaller GDP substrate. Within the TPS-d1 subfamily, the newly characterized (−)-linalool synthase from Norway spruce also clusters closely to these farnesene synthases from Norway spruce and loblolly pine. Similarly, in the TPS-b subfamily this cluster is comprised of *E,E*-α-farnesene synthase from *Malus* × *domestica* (AAO22848), the hemiterpene synthase isoprene synthase from *Populus alba* × *Populus tremula* (Miller et al., 2001), and geraniol synthase from *Cinnamomum tenuipilum* (CAD29734), and likely the geraniol synthase from *Ocimum basilicum* (Iijima et al., 2004; not shown). The presence of sesqui-TPS in these mono-TPS subfamilies (and a hemi-TPS in TPS-b) was supported by high bootstrap values and demonstrates a similar evolution of these TPS within both angiosperms and gymnosperms.

Another interesting feature of the gymnosperm and angiosperm TPS gene family is the presence of multiple clusters of limonene synthases (TPS-d1 and TPS-b), myrcene synthases (TPS-d1, TPS-b, and TPS-g) and linalool synthases (TPS-d1, TPS-b, TPS-f, and TPSg; Fig. 8). The phylogenetic distance between these TPS suggests multiple gene duplication events both in gymnosperms and angiosperms were followed by multiple events of convergent evolution to arrive at the same function. The ability of distantly related TPS of secondary metabolism to produce similar or identical products is a puzzling feature of the TPS gene family and again makes the case for experimental assessment of gene function over in silico comparative assignment of function.

DISCUSSION

In an effort to identify genes of direct and putative indirect terpenoid defenses in Norway spruce (Martin et al., 2002, 2003) and to provide a better foundation for phylogenetic analysis of the gymnosperm TPS family, we have cloned and functionally characterized a set of 9 new *TPS* genes from this conifer species. Along with the previously described (+)-3-carene synthase (Fäldt et al., 2003b), the 10 Norway spruce *TPS* genes represent one of the largest and functionally most diverse sets of characterized *TPS* from a single species. This suite of genes contains members encoding for mono-TPS, sesqui-TPS, and di-TPS each with a distinct product profile of terpenoids known to occur in Norway spruce resin defense and volatile terpenoid emissions (Martin et al., 2002, 2003). Among the *TPS* genes described here are two of biochemical functions that have not previously been described in any species, namely the multi-product sesqui-TPS, longifolene synthase, and the new di-TPS, PaTPS-Iso. In addition, we identified the first linalool synthase, a monoterpenol synthase, from a gymnosperm.

Sequence Relatedness of Norway Spruce TPS

As with TPS from grand fir (Bohlmann et al., 1999), the mono-TPS proteins from Norway showed in the range of 61% to 69% identity and 75% to 83% similarity among each other (Table II). Among the Norway spruce mono-TPS we discovered a gymnosperm linalool synthase that is only very distantly related in sequence to known angiosperm linalool synthases from *C. breweri* (14% I, 25% S; Dudareva et al., 1996) and Arabidopsis (23% I, 44% S; Chen et al., 2003; Fig. 8), but rather resembles conifer TPS that

form monoterpene hydrocarbons, suggesting that linalool synthase activity evolved independently in gymnosperms and angiosperms. Furthermore, since this enzyme is also able to synthesize small amounts of cyclic monoterpenes, the production of linalool may be the result of recent amino acid changes in a TPS derived from a regular mono-TPS. Interestingly, the newly identified Norway spruce *E,E*-α-farnesene synthase (PaTPS-Far), a sesqui-TPS, is also very similar to the conifer mono-TPS (60%–67% I, 72%–79% S); however, the lack of a transit peptide was the first indication for a sesqui-TPS that was confirmed by functional characterization of the expressed protein. The mono-TPS and *E,E*-α-farnesene synthase are much less similar to the other two Norway spruce sesqui-TPS, longifolene synthase and *E*-α-bisabolene synthase (28%–40% I, 42%–63% S) and to the two di-TPS (24%–25% I, 36%–38% S).

Surprisingly, the three Norway spruce sesqui-TPS described here fall into three distinct groups of the gymnosperm TPS-d subfamily (Fig. 7). *E,E*-α-farnesene synthase (PaTPS-Far) is closest in sequence to mono-TPS (TPS-d1). The multi-product longifolene synthase (PaTPS-Lon) is most closely related to the previously characterized grand fir γ-humulene synthase and δ-selinene synthase (Steele et al., 1998a) of the newly defined TPS-d2 group. The third spruce sesqui-TPS, *E*-α-bisabolene synthase (PaTPS-Bis), is most similar to its ortholog in grand fir and clusters closely with conifer di-TPS of the TPS-d3 group containing a 200- to 215-amino acid motif of yet unknown function. This cluster association pattern suggests that three types of sesqui-TPS evolved several times independently within the conifer TPS-d family from ancestors shared with di-TPS or mono-TPS. For instance, Norway spruce *E*-α-bisabolene synthase and *E,E*-α-farnesene synthase may have evolved from di-TPS and mono-TPS, respectively, after loss of exons coding for transit-peptides (Trapp and Croteau, 2001b) and specialization for FDP as substrate in the cytosol. Curiously, both *E*-α-bisabolene synthase and *E,E*-α-farnesene synthase catalyze relatively simple sesqui-TPS reactions, each yielding only a single acyclic or monocyclic product (Table III). These functions may arise more easily than the complex sesqui-TPS cyclization mechanisms. The more complex cyclization reactions, in conifers, all seem to come with very elaborate multiple-product profiles as seen with members of the TPS-d2 subgroup in Norway spruce (this study) and grand fir (Steele et al., 1998a).

The two Norway spruce di-TPS, PaTPS-LAS and PaTPS-Iso, are very similar (90% I, 94% S) to each other and much less similar to either sesqui-TPS (38%–43% I, 57%–63% S) or mono-TPS (24%–42% I, 36%–61% S). Remarkably, these two enzymes represent the two most similar TPS found in Norway spruce; however, their product profiles are entirely different (Fig. 5). It is also interesting to note that despite very similar amino acid sequences, PaTPS-Iso is a single-product enzyme, whereas PaTPS-LAS is a multi-product enzyme. Mechanistic differences between PaTPS-LAS and PaTPS-Iso most likely reside in a few amino acid differences within the active sites of these paralogous enzymes (see below).

Evolution of Gymnosperm TPS

The large TPS family of angiosperms has been divided into subfamilies based on sequence relatedness and functional assessment (Bohlmann et al., 1998b) and based on gene architecture (Trapp and Croteau, 2001b). Our phylogenetic analyses show that independent groups also exist within the gymnosperm TPS-d subfamily and that a further division based on these groups is justified. The inclusion of a large number of functionally characterized Norway spruce TPS, described here, as well as several gymnosperm TPS from the literature, enabled an expansion of the TPS-d tree into three groups based on sequence relatedness and biochemical classes. This analysis also indicated that some manifestation of biochemical functions of very similar di-TPS predated not only the separation of firs and spruces within the pine family, but in fact predated the separation of the Ginkgoales and the Coniferales. Since the Ginkgoales and the Coniferales evolved from a common progymnosperm ancestor (Chaw et al., 1997), this finding suggests that di-TPS of secondary metabolism existed prior to the division of these two orders. In this evolutionary model, we should expect to also find di-TPS of secondary metabolism within other gymnosperm orders, such as the Cycadales.

Our findings with enzymes of all three biochemical classes of TPS, mono-TPS, sesqui-TPS, and di-TPS, show that similar conifer TPS do group together and therefore support an evolutionary model that suggests specialization of many TPS biochemical functions prior to conifer speciation. A number of TPS homologs in the three genera Picea, Abies, and Pinus produce the same product profile, while in other homologous TPS, some sequence mutations have resulted in slightly modified product profiles of these enzymes. With the isolation and phylogenetic analysis of an increasing number of gymnosperm TPS it will become increasingly feasible to pinpoint sequence mutations that lead to functional diversification and to discriminate those from functionally silent amino acid substitutions.

Our new snapshot of the gymnosperm TPS-d family and comparison with the angiosperm TPS suggests that the TPS-d subfamily may have as many limbs as does the angiosperm tree of TPS (Fig. 8). When analyzing the gymnosperm TPS-d subfamily within the larger family of known plant TPS, several important features were observed. Gymnosperm di-TPS cluster closely to the di-TPS of primary gibberellic acid metabolism of the angiosperms (TPS-c and e) that all contain the ancestral approximately 200-amino acid di-TPS motif (Bohlmann et al., 1998b). This is similar to the result found when analyzing genomic TPS sequences (Trapp and Croteau, 2001b). However, the

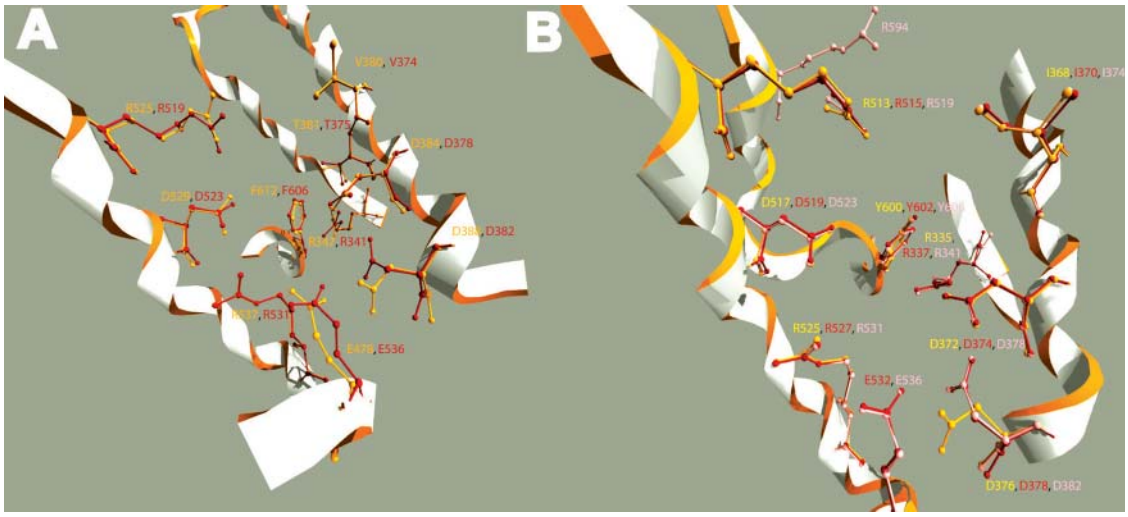

Figure 9. Comparative modeling of Norway spruce mono-TPS based on the crystal structure of bornyl diphosphate synthase from *S. officialis*. A, Models of Norway spruce myrcene synthase (orange) and grand fir myrcene synthase (red) show identical amino acids present in the active site of these two homologous enzymes. B, (−)-pinene synthase models from Norway spruce (orange), Sitka spruce (pink), and grand fir (red) show that these enzymes differ in the proximity of several amino acids with respect to the active site. Fewer amino acids seen in the Norway spruce model may help explain why this enzyme produces many products as opposed to only two products formed from the other (−)-pinene synthases.

gymnosperm di-TPS (TPS-d3) still group cohesively with the rest of the TPS-d subfamily and members that do not contain the ancestral approximately 200 amino acid motif. Therefore, one important missing link between the gymnosperm and angiosperm TPS families, and potentially a shared root of these families, remains to be found, possibly with future discovery of a gymnosperm di-TPS for gibberellic acid formation.

Compelling similarities between angiosperm (TPS-b) and the gymnosperm (TPS-d1) TPS were found regarding simple sesqui-TPS, which cluster closely with mono-TPS that produce oxygenated

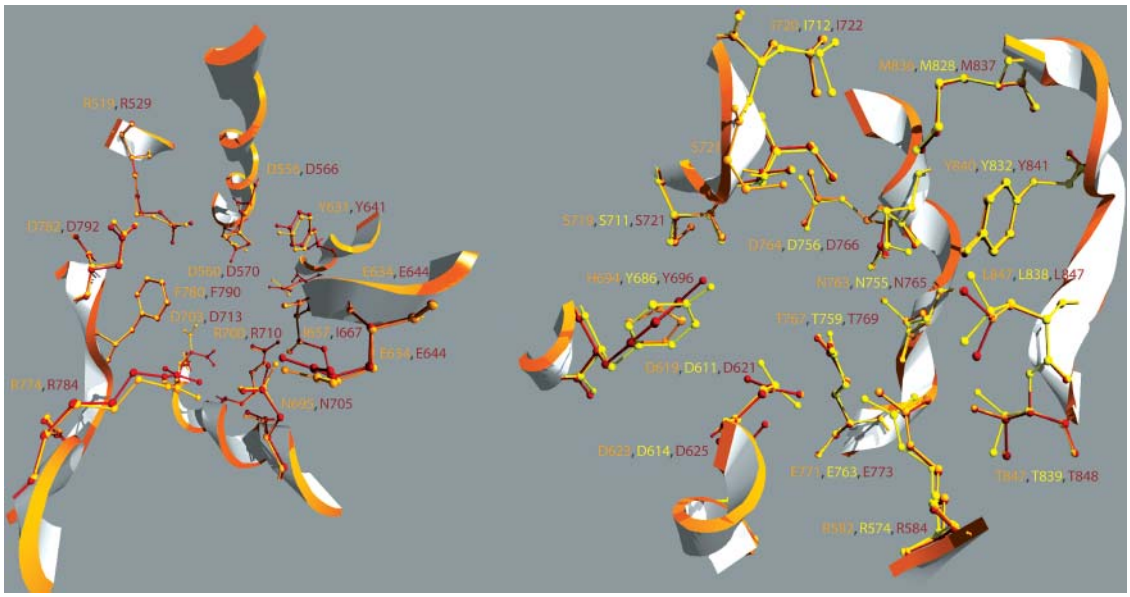

Figure 10. Comparative modeling of Norway spruce and grand fir sesqui-TPS and di-TPS. A, E-α-bisabolene synthases from Norway spruce (orange) and grand fir (red) demonstrate the same amino acids model in the active site of these homologs. B, Models of PaTPS-Iso (orange) and PaTPS-LAS (yellow) from Norway spruce as compared to abietadiene synthase AgAS (red) from grand fir demonstrate two main amino acid differences: H964_PaTPS-Iso versus Y686_PaTPS-LAS and Y696_AgAS; and S721_PaTPS-Iso versus A713_PaTPS-LAS (not shown) and A721_AgAS (not shown) between these three related proteins. Differences are unique to the functionally distinct PaTPS-Iso.

monoterpenes and with hemi-TPS (angiosperms). This demonstrates simple sesqui-TPS have evolved from mono-TPS in both gymnosperms and angiosperms. Likewise, the presence of multiple rather distantly related limonene synthases, myrcene synthases, and linalool synthases provide evidence that these TPS evolved more than once in angiosperms and gymnosperms.

Comparative Modeling of Norway Spruce TPS Active Site Structures

The cloning, functional characterization and phylogenetic analysis of a family of Norway spruce mono-, sesqui-, and diTPS desribed in this study identified several interesting candidate enzymes for future structure-function analysis of TPS catalysis. In the absence of any authentic three-dimensional structure for a gymnopserm TPS, comparative protein modeling against more distant angiosperm TPS can reveal valuable information about amino acid positions in conserved proteins. The x-ray structures of 5-*epi*-aristolochene synthase (NtEAS) from *Nicotiana tabacum* (Starks et al., 1997) as well as bornyl diphosphate synthase (SoBDS) from *Salvia officialis* (Whittington et al., 2002) provide the necessary templates for the modeling of conifer TPS (Little and Croteau, 2002; Peters and Croteau, 2002; Phillips et al., 2003). Through comparative modeling we hope to pinpoint amino acids involved in catalysis and provide a starting point for future site directed mutagenesis of selected conifer TPS.

Initially, we selected two closely related myrcene synthases one from Norway spruce (PaTPS-Myr) and one from grand fir (AgMyr) to test the feasibility of comparative modeling of mono-TPS from two different species against the angiosperm SoBDS (Whittington et al., 2002). We found that these two conifer mono-TPS do have the same amino acids present in their modeled active sites (Fig. 9A). Similarly, pinene synthases from Norway spruce (PaTPS-Pin), Sitka spruce (PsPin), and grand fir (AgPin) were also modeled. These mono-TPS are similar but all produce slightly different product profiles (Supplemental Table II). Models of the three pinene synthase enzymes were very similar (Fig. 9B). Several amino acids seen in the AgPin (E532) and PsPin (E536 and R594) active site models were not seen within the PaTPS-Pin model in a 3-Å snapshot. We also chose to model two conifer sesqui-TPS, *E*-α-bisabolene synthases, from Norway spruce (PaTPS-Bis) and grand fir (AgBis) based on NtEAS (Starks et al., 1997). Comparative modeling placed the same amino acids within the active site of both of these enzymes (Fig. 10A). These comparative models have painted a picture of some of the amino acids involved in the active sites of these TPS for future analysis and provided some level of confidence for modeling of conifer mono-TPS and sesqui-TPS, two biochemical classes of TPS for which three-dimensional structures exist with angiosperm TPS.

Recent work on abietadiene synthase (AgAS) from grand fir (Ravn et al., 2000, 2002; Peters et al., 2000, 2001, 2003) has dissected the mechanism of this bifunctional conifer di-TPS. AgAS is very similar to Norway spruce PaTPS-LAS and PaTPS-Iso, thereby making the recent findings relevant to the mechanisms of these two enzymes as well. While PaTPS-LAS and AgAS have nearly identical qualitative and quantitative muliple-product profiles, the single product of PaTPS-Iso is distinctively different. Yet PaTPS-Iso is overall more similar in amino acid sequence to PaTPS-LAS than PaTPS-LAS and AgAs are to each other.

Modeling the active sites of PaTPS-LAS, PaTPS-Iso, and AgAS on NtEAS (Starks et al., 1997) revealed only two amino acid differences within this region (Fig. 10B). One such difference is located within the conserved TPS active site of the PaTPS-LAS and PaTPS-Iso models and is defined by an overlapping amino acid difference at Y686_PaTPS-LAS, Y696_AgAS, and H694_PaTPS-Iso. The only other difference in these models is the presence of S721_PaTPS-Iso. In both PaTPS-LAS and AgAS, this position is occupied by A713 and A723, respectively (not shown). Levopimaradiene synthase from *G. biloba* (Schepmann et al., 2001), also capable of producing an abietane skeleton, has a Y700 and an A727 at these positions. These results suggest that comparative modeling of conifer di-TPS against a distant angiopserm sesqui-TPS could pinpoint the relevant amino acid substitutions in otherwise highly conserved pairs of di-TPS. These modeling studies will direct future experiments that will analyze the involvement of these amino acids in catalysis.

CONCLUSION

In summary, the family of 10 functionally characterized *TPS* genes from Norway spruce is critical to quality and quantity and the structural diversity of terpenoid defenses in this species. The isolation of these genes has provided a means to better understand some of the phylogenetic relationships that exist in the TPS-d subfamily and the TPS family. Comparative modeling can provide further information beyond simple sequence alignments toward the identification of specific amino acids within the active sites of conifer TPS enzymes and will guide future structure-function analysis. Furthermore, the suite of 10 spruce *TPS* genes has provided the tools for future transcript profiling of *mono-TPS*, *sesqui-TPS*, and *di-TPS* in stem tissues and foliage in response to MeJA-treatment and in response to attack by stem boring insects (*Pissodes strobi*) in species of spruce (B. Miller, L.L. Madilao, S. Ralph, and J. Bohlmann, unpublished data). These studies will comprehensively extend on previous TPS gene expression analysis in the spruce system (Byun McKay et al., 2003; Fäldt et al., 2003b) and show that *TPS* gene expression profiles closely match quantitative and

qualitative changes in terpenoid metabolites in stems and terpenoid volatile emissions in foliage.

MATERIALS AND METHODS

Materials

Norway spruce (*Picea abies*) seedlings (clone 244–932) used for RNA isolations and cDNA preparation were previously described (Fäldt et al., 2003b). RNA for RACE was prepared from 2-year-old Norway spruce seedlings grown under natural environmental conditions on the University of British Columbia campus. Substrates (2 mg/mL), GDP and FDP, were from Echelon Bioscience (Salt Lake City). GGDP was from Sigma-Aldrich Canada. (Oakville, Ontario). Unless otherwise mentioned, all other reagents and solvents were from Fischer Scientific (Pittsburgh), Sigma-Aldrich, or VWR International (West Chester, PA). All standards used in GC analysis were of the highest purity available.

General Molecular Biology Procedures

Routine PCR reactions were performed in volumes of 50 μL containing 20 mM Tris/HCl (pH 8.4), 50 mM KCl, 5 mM $MgCl_2$, 200 μM of each dNTP, 0.1 μM of a forward primer and a reverse primer, one unit of Taq Polymerase (Invitrogen, Carlsbad, CA), and 2 ng of template DNA. PCR amplifications for subcloning of *TPS* cDNAs were performed in volumes of 50 μL containing 20 mM Tris/HCl (pH 8.8), 10 mM KCl, 10 mM $(NH_4)_2SO_4$, 2 mM $MgSO_4$, 0.1% Triton X-100, 10 μg BSA, 200 μM of each dNTP, 0.1 μM of each primer, and 2.5 units of high fidelity Turbo *Pfu* polymerase (Stratagene, La Jolla, CA). All PCR reactions were performed using MJ PTC100-thermocyclers (Waltham, MA). QIAquick Gel Extraction kits (Qiagen, Valencia, CA) were used to extract DNA from agarose gels and plasmid DNA was isolated using QIAprep Spin Miniprep kits (Qiagen). DNA was sequenced using DyeDeoxy Terminator Cycle Sequencing (Applied Biosystems, Foster City, CA) at the Nucleic Acids Protein Services Unit (UBC, Vancouver). For hybridization experiments, DNA was labeled with [α-^{32}P]dCTP (Easytide, Perkin Elmer, Foster City, CA) using the Rediprime II random prime labeling system (Amersham-Pharmacia Biotech; Buckinghamshire, UK). Cloning vectors included the pCR-Blunt vector (Invitrogen), the pCR2.1-TOPO (Invitrogen), the pET100/D-TOPO and pET101/D-TOPO directional expression vectors (Invitrogen), and pSBET (Schenk et al., 1995).

cDNA Library Screening

A Norway spruce young shoot λ-ZAP phage library was screened for mono-TPS and sesqui-TPS exactly as described in Fäldt et al. (2003b). For isolation of di-TPS and sesqui-TPS containing the 210 amino acid element (Bohlmann et al., 1998a, 1998b), a portion of this region was amplified using the primers ASDL1 and ASDL2 (Supplemental Table III) and 2 μL of the λ-ZAPII cDNA library as a template. The resulting 540-bp amplicon was cloned into pCR2.1-TOPO (Invitrogen) and transformed into *Escherichia coli* TOP10F′ cells (Invitrogen). After single colony isolation and sequence verification from plasmid DNA, the 540-bp probe was amplified by PCR, purified by agarose gel electrophoresis, labeled with Rediprime procedure, and used to screen by filter hybridization the Norway spruce cDNA library as described (Fäldt et al., 2003b). A total of 290 primary positive phage plaques yielded 72 pBluescript SK(−) phagemids isolated with the mono-TPS probe (Fäldt et al., 2003b) and 8 pBluescript SK(−) phagemids isolated with the di-TPS/sesqui-TPS probe after secondary screening. Phagemids were in vivo excised in *E. coli* XL1 Blue MRF′ and transformed into *E. coli* SOLR. Plasmid DNA was prepared from individual transformants, inserts sequenced, and compared to the GenBank database using blastx (www.ncbi.nlm.nih.gov/BLAST/). Bluescript plasmids identified as *PaTPS-Far*, *PaTPS-Lin*, *PaTPS-Pin*, and *PaTPS-Bis* showed high sequence similarities to mono-TPS and sesqui-TPS and were subcloned into expression vectors. Sequences of partial plasmid clones similar to conifer di-TPS were used to design primers for RACE.

RACE

Norway spruce seedlings were sprayed with 0.1 mM MeJA in a 2% (v/v) Tween 20/water solution. Trees were harvested 2 d, 4 d, and 6 d after treatment. Samples of needles were combined over the three collection days and RNA extracted according to the method by S.X. Wang et al. (2000). The FirstChoice RLM-RACE kit (Ambion, Austin, TX) was used to generate both 5′- and 3′-RACE PCR templates. Nested 5′-PCR reactions were carried out with 2 μL of the RACE template, 400 μM each dNTP, 200 mM Tris-HCl (pH 8.8), 20 mM $MgSO_4$, 100 mM KCl, 100 mM $(NH_4)_2SO_4$, 1% Triton X-100, 1 mg/mL nuclease-free bovine serum albumin, 3.125 units of PfuTrubo, and 0.4 μM each vector based forward primer and sequence based reverse primer. These conditions were held constant for all RACE PCR reactions. To isolate full-length di-TPS cDNAs, two sequential nested PCR reactions were performed using the outer Rc3-5c and inner Rc3-5a (Supplemental Table III) di-TPS primers. Products of this PCR reaction were cloned into the pCR-Blunt vector (Invitrogen) and transformed into *E. coli* TOP10F′ cells (Invitrogen). Single colonies were isolated, plasmid DNA was sequenced and the resulting sequences were organized into contigs. Using the sequence information, RCAAS5-31 outer and RCAAS5-32 (Supplemental Table III) inner forward primers were designed for 3′-RLM-RACE PCR. Using cDNA-based forward primers and vector-based reverse primers sequential nested PCR reactions were performed. Products were cloned into pCR-Blunt and sequenced as before. RACE methods produced overlapping sequences to discern the entire reading frame of di-TPS-like cDNA. Using this information, Pa24_05a forward and Pa2_3b reverse primers were designed and the 3′ RLM-RACE template was used to amplify the full-length cDNA. Products of this reaction were cloned into the pCR-Blunt vector (Invitrogen) and sequenced. From this one template two different full-length cDNAs (*PaTPS-LAS* and *PaTPS-Iso*) were isolated. Primers designed to target mono-TPS genes were used to amplify full-length genes from the 3′ RLM-RACE template. The primers 743_5′a and 743_3′stp were used in a primary reaction and 743_M5petTP and 743_3′stp in a secondary reaction, yielding the full-length cDNA, *PaTPS-Lim*. This PCR product was cloned directly into the pET101 (Invitrogen) expression vector and transformed into Top10 F′cells.

Subcloning of TPS cDNAs for Expression in *E. coli*

Plastidial targeting peptides are present in monoterpene and diterpene synthases and delineated by the RRX_8W motif (Williams et al., 1998; Bohlmann et al., 1998b). For several TPS, deletion of plastidial targeting peptide was shown to improve expression of functional TPS enzymes (Williams et al., 1998; Bohlmann et al., 1999). For most *mono-TPS* and *di-TPS* cDNAs described here, two constructs were made, one with and one without the transit peptide. For subcloning, *TPS* cDNAs were amplified by PCR using high fidelity Turbo *Pfu* polymerase (Stratagene) with 2 μL of the amplified λ-ZAPII phage library (Fäldt et al., 2003b) or 2 ng of plasmid DNA. For *PaTPS-Lin* the primer combinations cl39Pe5a and cl39Pet3a or cl39Pet5b and cl39Pet3a were used for full-length and truncated constructs, respectively. Amplification of *PaTPS-Far* utilized the primer combination 5′71RR_pET and 3′71_BAM3. Primers for *PaTPS-Bis* were pET100pa03F100 and pET100pa03aR2521. PCR products were cloned into the pET100/D-TOPO expression vector. *PaTPS-Lon* was amplified using λ-ZAPII phage library as template and primers designed for GenBank accession number AF369920, 486_M5pet and 486_3′stp. PCR product cloned into pET100/D-TOPO vector. According to Peters et al. (2000), the transit peptide of *A. grandis* abietadiene synthase is cleaved close to Leu-79 since this truncation allows maximum enzyme activity. Similar to the *A. grandis* protein, *PaTPS-LAS* and *PaTPS-Iso* also have Leus at amino acids 72 and 80, respectively. cDNA inserts of *PaTPS-LAS* and *PaTPS-Iso* in the pCR-Blunt vector were amplified by PCR to generate full-length and truncated versions for subcloning into the pET100/D-TOPO vector. Since these two genes are highly similar, it was possible to use the same 3′ primer, Pa2_3b, for amplification. The 5′ primer for full-length constructs of both *PaTPS-LAS* and *PaTPS-Iso* was 245petM and for truncated constructs, Pa24_pet5b and 25_L79_pet, were used for *PaTPS-LAS* and *PaTPS-Iso*, respectively.

The *PaTPS-Pin* cDNA insert of plasmid pBluescript- *PaTPS-Pin* was subcloned into pSBET expression vector. Introduction of *Nde*I and *Bam*HI ends was accomplished by the sticky end PCR cloning strategy (Zeng, 1998). Primer combinations for insert PCR amplification were JF104NDE1_103 and JF104BAM3_105 as well as JF104NDE2_104 and JF104BAM4_106. The two amplified segments of *PaTPS-Pin* were denatured (2 min at 95°C), combined, and reannealed (30 min at room temperature) to create sticky ends (*Nde*I and *Bam*HI), and subsequently ligated into *NdeI*/*Bam*HI-digested pSBETa as described by Bohlmann et al. (1999). Restriction endonuclease sites, *Kpn*I (5′ end) and *Pst*I (3′ end), were introduced into the *PaTPS-Myr* cDNA insert of plasmid *PaTPS-Myr* by PCR amplification using the combination of JB016RE_152 and JB016RE_153 primers. PCR product of *PaTPS-Myr* was

digested with *Kpn*I and *Pst*I, gel purified, and ligated into *Kpn*I/*Pst*I-digested pQE50. In all cases, the recombinant expression plasmids were rescued and maintained in *E. coli* TOP10 F′ cells, analyzed by PCR using insert and/or vector based primers, and sequenced. For expression, plasmids were transformed into *E. coli* BL21-CodonPlus (Stratagene) cells for the pET and pQE constructs or into BL21(DE3) Star (Invitrogen) cells for pSBET constructs.

Expression of TPS in *E. coli* and Enzyme Assays

Functional expression and enzyme assays have been previously described in Fäldt et al. (2003b). Briefly, cultures were grown in 5 mL Luria-Bertani and the appropriate antibiotic (either ampicillin or kanamycin) induced with 200 μM isopropylthio-β-galactoside, centrifuged, and resuspended in mono-TPS buffer (25 mM HEPES, pH 7.2, 100 mM KCl, 10 mM $MnCl_2$, 10% glycerol, 5 mM dithiothreitol [DTT]), sesqui-TPS buffer (25 mM HEPES, pH 7.3, 10 mM $MgCl_2$, 10% glycerol, 10 mM DTT), or di-TPS buffer (25 mM HEPES, pH 7.2, 100 mM KCl, 10 mM $MgCl_2$, 10 μM $MnCl_2$, 5% glycerol, 5 mM DTT). Cells were homogenized by ultrasound treatment, extracts were cleared by centrifugation, and the supernatant was removed and assayed for TPS activity with GDP, FDP, or GGDP. Substrate concentrations for assays were used at a final concentration of 137 μM GDP, 92.3 μM FDP, and 40 μM GGDP. A 1-mL pentane overlay was used to trap terpene products during the assays and in all cases, after incubation at 30°C for 1 h, the reaction mixture was further extracted with pentane (3 × 1 mL). The combined pentane fraction was extracted with water or purified over silica/$MgSO_4$ as described (Lewinsohn et al., 1991b; Lafever et al., 1994; Bohlmann et al., 1997; Peters et al., 2000; Martin et al., 2002; Fäldt et al., 2003b). Pentane extracts were then evaporated to 50 to 100 μL and 1 or 2 μL were injected and analyzed by GC-FID and GC-MS as previously described (Martin et al., 2002, 2003; Fäldt et al., 2003b). Controls for product formation independent of any cDNA were performed using extracts of *E. coli* BL21-CodonPlus (DE3) transformed with either plasmid without the insert.

Terpenoid Product Identification by GC-MS

Monoterpenes were identified and quantified by GC-MS analysis on an Agilent 6890 Series GC system coupled to an Agilent 5973 Network Mass Selective Detector (70 eV) using a DB-WAX (J&W Scientific, Palo Alto, CA) capillary column (0.25 mm i.d. × 30 m with 0.25-μm film) or an HP-5 (Agilent Technologies, Palo Alto, CA) capillary column (0.25 mm i.d. × 30 m with 0.25-μm film). An injector was used at 200°C and a column flow of 1 mL He min^{-1}. Using the DB-WAX column, the following temperature program was used: the initial temperature of 40°C (4-min hold) was increased to 150°C at 4°C min^{-1} followed by a 20°C min^{-1} ramp until 230°C (5-min hold). The temperature program for the HP-5 column began at 40°C (2-min hold) with an initial ramp of 3°C min^{-1} until 160°C followed by a 10°C min^{-1} increase to 200°C. A final increase of 20°C min^{-1} until 300°C (3-min hold) completed the program. A Cyclodex B (permethylated β-cyclodextrin in DB-1701, J&W Scientific) capillary column (0.25 mm i.d. × 30 m with 0.25-μm film) was used for separation of enantiomers of *PaTPS-Pin* monoterpenoid products. For separation enantiomers of α-pinene and β-pinene, sabinene, limonene, and β-phellandrene, the following temperature program was used: initial temperature was 55°C (1-min hold), which was then increased to 100°C at 1°C min^{-1} followed by a 10°C min^{-1} ramp until 230°C (10-min hold). Analysis of enantiomers products of *PaTPS-Lin* and *PaTPS-Lim* were performed with a Cyclosil B column [30% hepatkis (2,3-di-*O*-methyl-6-*O*-t-butyl dimethylsilyl)-β-cyclodextrin in DB-1701; Agilent Technologies] capillary column (0.25 mm i.d. × 30 m with 0.25-μm film). To separate linalool enantiomers, the following program was used on an Agilent 6890 GC fitted with an FID: 60°C (6-min hold) initial temperature then 3°C min^{-1} until 180°C followed by 20°C min^{-1} to 220°C (5-min hold). For the separation of limonene, α-pinene, and β-pinene enantiomers, the program was as follows: 60°C (6-min hold) initial temperature then 3°C min^{-1} until 180°C followed by 20°C min^{-1} to 220°C (5-min hold). Sesquiterpenes and diterpenes were initially separated on an HP-5 capillary column (0.25 mm i.d. × 30 m with 0.25-μm film). Sesquiterpenes were separated with the following program of 40°C (1-min hold), then 5°C min^{-1} until 180°C, followed by 20°C min^{-1} until 300°C (1-min hold). Diterpenes were separated on the HP-5 column using the following program: 40°C (1-min hold), then 7.5°C min^{-1} until 210°C followed by 3°C min^{-1} until 280°C (5-min hold). Additional product analysis and confirmation of results obtained with the HP5 column were performed on the DB-WAX column (see above). The program used with the DB-WAX column was 40°C (3-min hold), increased to 250°C at 5°C min^{-1} (5-min hold). All terpenoid products were identified using authentic standards and/or library matches. Libraries utilized include the Wiley library and The Identification of Essential Oil Components by Gas Chromatography/Quadrapole Mass Spectroscopy library (Adams, 2002). Further confirmation of product ID was done by separation on more than one type of column combined with Kovat's index information (Adams, 2002).

Sequence and Phylogenetic Analyses

Predictions for pI were made using the entire open reading frame within EditSeq 5.00 (DNASTAR). Amino acid alignments were made with ClustalX (www-igbmc.u-strasbg.fr/BioInfo/) and GeneDoc (www.psc.edu/biomed/genedoc/). Since transit peptides are not well conserved, these were truncated from mono-TPS and di-TPS prior to analyzing TPS for phylogenetic relationships. The sequences were then aligned using Dialign (Morgenstern et al., 1998; available through MAGI at the HGMP www.hgmp.mrc.ac.uk), a program capable of finding local similarities in divergent sequences. Multiple sequence alignments were then hand corrected using GeneDoc (www.psc.edu/biomed/genedoc/). Maximum likelihood analyses using the data sets for the TPS-d subfamily as well as for the entire TPS family were analyzed by Phyml (Guindon and Gascuel, 2003) using the JTT (Jones et al., 1992) amino acid substitution matrix. The proportion of invariable sites as well as the alpha shape parameter was estimated by Phyml. Trees were generated using BIONJ (Gascuel, 1997), a modified neighbor joining algorithm. SEQBOOT of the Phylip 3.6 package (Felsenstein, 1993; evolution.genetics.washington.edu/phylip.html) was used to generate 100 bootstrap replicates. These were then analyzed by Phyml using the previously estimated parameters. CONSENSE, also from the Phylip 3.6 package (evolution.genetics.washington.edu/phylip.html), was used to generate a consensus tree. Distance analyses were completed on 1,000 bootstrap replicated data sets (SEQBOOT) using PROTDIST and Dayhoff PAM 001 matrix (Dayhoff, 1979). NEIGHBOR and CONSENSE were used to generate the neighbor-joining consensus tree. All distance analyses were completed using programs of the Phylip 3.5c package (Felsenstein, 1993) available through UBiC Resources and Support at the UBC Bioinformatics Centre, University of British Columbia (www.bioinformatics.ubc.ca). Treeview (Page, 1996) was used to visualize all trees. Bootstrap values above 50% for both the maximum likelihood analysis and the distance analysis were added to the maximum likelihood tree calculated from the original data set.

Comparative Modeling

Comparative modeling of TPS was accomplished using the DeepView/Swiss-PDB Viewer and the SWISS-PROT Modeler (Guex and Peitsch, 1997). Energy was minimized by GROMOS96 (iqc.ethz.ch/grooms; Van Gunsteren et al., 1996) executed within the DeepView/Swiss-PDB Viewer interface. Snapshot views were examined by choosing amino acids with 3.25 Å (mono-TPS), 3.75 Å (sesqui-TPS), or 4.0 Å (di-TPS) of the substrate analog and Mg^{2+} ions from either bornyl diphosphate synthase from *S. officialis* (mono-TPS) or 5-*epi*-aristolochene synthase from *N. tabacum* (sesqui-TPS and di-TPS). Views were rendered using PovRay 3.5 (www.povray.org).

Distribution of Materials

Upon request, all novel materials described in this publication will be made available in a timely manner under standard material transfer agreements for noncommercial research purposes.

Sequence data from this article have been deposited with the EMBL/GenBank data libraries under accession numbers AY473619 to AY473627.

ACKNOWLEDGMENTS

We thank Dr. Robert Coates and Yinghua (Alice) Jin for diterpene standards, Dr. Rodney Croteau for the *A. grandis* abietadiene

synthase clone, and Drs. Reuben Peters and Patrick Keeling for helpful advice.

Received March 3, 2004; returned for revision March 25, 2004; accepted March 25, 2004.

LITERATURE CITED

Adams RP (2002) The Identification of Essential Oil Components by Gas Chromatography/Quadrupole Mass Spectroscopy. Allured Publishing, Carol Stream, IL

Alfaro RI, Borden JH, King JN, Tomlin ES, McIntosh RL, Bohlmann J (2002) Mechanisms of resistance in conifers against shoot infesting insects. *In* MR Wagner, KM Clancy, F Lieutier, TD Paine, eds, Mechanisms and Deployment of Resistance in Trees to Insects. Kluwer Academic Press, Dordrecht, The Netherlands, pp 101–126

Arimura G, Huber DPW, Bohlmann J (2004) Forest tent caterpillars (*Malacosoma disstria*) induce local and systemic diurnal emissions of terpenoid volatiles in hybrid poplar (*Populus trichocarpa x deltoides*): cDNA cloning, functional characterization, and patterns of gene expression of (−)-germacrene D synthase, *PtdTPS1*. Plant J **37:** 603–616

Aubourg S, Lecharny A, Bohlmann J (2002) Genomic analysis of the terpenoid synthase (*AtTPS*) gene family of *Arabidopsis thaliana*. Mol Genet Genomics **267:** 730–745

Bohlmann J, Crock J, Jetter R, Croteau R (1998a) Terpenoid-based defenses in conifers: cDNA cloning, characterization, and functional expression of wound-inducible (*E*)-α-bisabolene synthase from grand fir (*Abies grandis*). Proc Natl Acad Sci USA **95:** 6756–6761

Bohlmann J, Croteau R (1999) Diversity and variability of terpenoid defenses in conifers: molecular genetics, biochemistry and evolution of the terpene synthase gene family in grand fir (*Abies grandis*). *In* DJ Chadwick, JA Goode, eds, Insect Plant Interactions and Induced Plant Defense. John Wiley and Sons, Chichester, UK, pp 132–146

Bohlmann J, Meyer-Gauen G, Croteau R (1998b) Plant terpenoid synthases: Molecular biology and phylogenetic analysis. Proc Natl Acad Sci USA **95:** 4126–4133

Bohlmann J, Phillips M, Ramachandiran V, Katoh S, Croteau R (1999) cDNA cloning, characterization, and functional expression of four new monoterpene synthase members of the *Tpsd* gene family from grand fir (*Abies grandis*). Arch Biochem Biophys **368:** 232–243

Bohlmann J, Steele CL, Croteau R (1997) Monoterpene synthases from grand fir (*Abies grandis*). cDNA isolation, characterization, and functional expression of myrcene synthase, (−)-(*4S*)-limonene synthase, and (−)-(*1S,5S*)-pinene synthase. J Biol Chem **272:** 21784–21792

Byun McKay A, Hunter W, Goddard K, Wang S, Martin D, Bohlmann J, Plant A (2003) Insect attack and wounding induce traumatic resin duct development and gene expression of (−)-pinene synthase in Sitka spruce. Plant Physiol **133:** 368–378

Cane DE (1999) Sesquiterpene biosynthesis: cyclization mechanisms. *In* DE Cane, ed, Comprehensive Natural Products Chemistry: Isoprenoids, Including Carotenoids and Steroids, Vol 2. Pergamon Press, Oxford, pp 155–200

Chaw S-M, Zharkikh A, Sung H-M, Lau T-C, Li W-H (1997) Molecular phylogeny of extant gymnosperms and seed plant evolution: analysis of nuclear 18S rRNA sequences. Mol Biol Evol **14:** 56–68

Chen F, Tholl D, D'Auria JC, Farooq A, Pichersky E, Gershenzon J (2003) Biosynthesis and emission of terpenoid volatiles from *Arabidopsis* flowers. Plant Cell **15:** 481–494

Croteau R, Kutchan T, Lewis N (2000) Natural products (secondary metabolism). *In* BB Buchanan, W Gruissem, RL Jones, eds, Biochemistry and Molecular Biology of Plants. American Society of Plant Biologists, Rockville, MD, pp 1250–1318

Davis EM, Croteau R (2000) Cyclization enzymes in the biosynthesis of monoterpenes, sesquiterpenes, and diterpenes. Top Curr Chem **209:** 53–95

Dayhoff MO (1979) Atlas of Protein Sequence and Structure, Vol 5, Supplement 3. National Biomedical Research Foundation, Washington, DC

Dudareva N, Cseke L, Blanc VM, Pichersky E (1996) Evolution of floral scent in *Clarkia*: novel patterns of S-linalool synthase gene expression in the *C. breweri* flower. Plant Cell **8:** 1137–1148

Dudareva N, Martin D, Kish CM, Kolosova N, Gorenstein N, Fäldt J, Miller B, Bohlmann J (2003) (E)-β-Ocimene and myrcene synthase genes of floral scent biosynthesis in snapdragon: function and expression of three terpene synthase genes of a new terpene synthase subfamily. Plant Cell **15:** 1227–1241

Fäldt J, Arimura G, Gershenzon J, Takabayashi J, Bohlmann J (2003a) Functional identification of AtTPS03 as (*E*)-β-ocimene synthase: a monoterpene synthase catalyzing jasmonate- and wound-induced volatile formation in *Arabidopsis thaliana*. Planta **216:** 745–751

Fäldt J, Martin D, Miller B, Rawat S, Bohlmann J (2003b) Traumatic resin defense in Norway spruce (*Picea abies*): Methyl jasmonate-induced terpene synthase gene expression, and cDNA cloning and functional characterization of (+)-3-carene synthase. Plant Mol Biol **51:** 119–133

Felsenstein J (1993) PHYLIP (Phylogeny Inference Package) Version 3.5c. Distributed by the author. Department of Genetics, University of Washington, Seattle

Franceschi VR, Krekling T, Christiansen E (2002) Application of methyl jasmonate on *Picea abies* (Pinaceae) stems induces defense-related responses in phloem and xylem. Am J Bot **89:** 578–586

Gascuel O (1997) BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. Mol Biol Evol **14:** 685–695

Guex N, Peitsch MC (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis **18:** 2714–2723

Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol **52:** 696–704

Iijima Y, Gang DR, Fridman E, Lewinsohn E, Pichersky E (2004) Characterization of geraniol synthase from the peltate glands of sweet basil. Plant Physiol **134:** 370–379

Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. Comput Appl Biosci **8:** 275–282

Kessler A, Baldwin IT (2002) Plant responses to insect herbivory: the emerging molecular analysis. Annu Rev Plant Biol **53:** 299–328

LaFever RE, Stofer-Vogel B, Croteau R (1994) Diterpenoid resin acid biosynthesis in conifers – enzymatic cyclization of geranylgeranyl pyrophosphate to abietadiene, the precursor of abietic acid. Arch Biochem Biophys **313:** 139–149

Lesburg CA, Zhai G, Cane DE, Christianson DW (1997) Crystal structure of pentalenene synthase: mechanistic insights on terpenoid cyclization reactions in biology. Science **277:** 1820–1824

Lewinsohn E, Gijzen M, Savage TJ, Croteau R (1991b) Defense-mechanisms of conifers - Relationship of monoterpene cyclase activity to anatomical specialization and oleoresin monoterpene content. Plant Physiol **96:** 38–43

Little DB, Croteau RB (2002) Alteration of product formation by directed mutagenesis and truncation of the multiple-product sesquiterpene synthases delta-selinene synthase and gamma-humulene synthase. Arch Biochem Biophys **402:** 120–135

Martin D, Tholl D, Gershenzon J, Bohlmann J (2002) Methyl jasmonate induces traumatic resin ducts, terpenoid resin biosynthesis, and terpenoid accumulation in developing xylem of Norway spruce stems. Plant Physiol **129:** 1003–1018

Martin DM, Gershenzon J, Bohlmann J (2003) Induction of volatile terpene biosynthesis and diurnal emission by methyl jasmonate in foliage of Norway spruce (*Picea abies*). Plant Physiol **132:** 1586–1599

Miller B, Oschinski C, Zimmer W (2001) First isolation of an isoprene synthase gene from poplar and successful expression of the gene in *Escherichia coli*. Planta **213:** 483–487

Morgenstern B, Frech K, Dress A, Werner T (1998) DIALIGN: finding local similarities by multiple sequence alignment. Bioinformatics **14:** 290–294

Page RDM (1996) TREEVIEW: an application to display phylogenetic trees on personal computers. Comput Appl Biosci **12:** 357–358

Peters RJ, Carter OA, Zhang Y, Matthews BW, Croteau RB (2003) Bifunctional abietadiene synthase: mutual structural dependence of the active sites for protonation-initiated and ionization-initiated cyclizations. Biochemistry **42:** 2700–2707

Peters RJ, Croteau RB (2002) Abietadiene synthase catalysis: conserved residues involved in protonation-initiated cyclization of geranyl-

geranyl diphosphate to (+)-copalyl diphosphate. Biochemistry **41:** 1836–1842

Peters RJ, Flory JE, Jetter R, Ravn MM, Lee HJ, Coates RM, Croteau RB (2000) Abietadiene synthase from grand fir (*Abies grandis*): characterization and mechanism of action of the ''pseudomature'' recombinant enzyme. Biochemistry **39:** 15592–15602

Peters RJ, Ravn MM, Coates RM, Croteau RB (2001) Bifunctional abietadiene synthase: free diffusive transfer of the (+)-copalyl diphosphate intermediate between two distinct active sites. J Am Chem Soc **123:** 8974–8978

Phillips MA, Croteau RB (1999) Resin-based defenses in conifers. Trends Plant Sci **4:** 184–190

Phillips MA, Savage TJ, Croteau R (1999) Monoterpene synthases of loblolly pine (*Pinus taeda*) produce pinene isomers and enantiomers. Arch Biochem Biophys **372:** 197–204

Phillips MA, Wildung MR, Williams DC, Hyatt DC, Croteau R (2003) cDNA isolation, functional expression, and characterization of (+)-α-pinene synthase and (−)-α-pinene synthase from loblolly pine (*Pinus taeda*): stereocontrol in pinene biosynthesis. Arch Biochem Biophys **411:** 267–276

Ravn MM, Coates RM, Flory JE, Peters RJ, Croteau R (2000) Stereochemistry of the cyclization-rearrangement of (+)-copalyl diphosphate to (−)-abietadiene catalyzed by recombinant abietadiene synthase from *Abies grandis*. Org Lett **2:** 573–576

Ravn MM, Peters RJ, Coates RM, Croteau R (2002) Mechanism of abietadiene synthase catalysis: stereochemistry and stabilization of the cryptic pimarenyl carbocation intermediates. J Am Chem Soc **124:** 6998–7006

Schenk PM, Baumann S, Mattes R, Steinbiss HH (1995) Improved high-level expression system for eukaryotic genes in *Escherichia coli* using T7 RNA-polymerase and rare (Arg) tRNAs. Biotechniques **19:** 196–198

Schepmann HG, Pang J, Matsuda SP (2001) Cloning and characterization of *Ginkgo biloba* levopimaradiene synthase which catalyzes the first committed step in ginkgolide biosynthesis. Arch Biochem Biophys **392:** 263–269

Schnee C, Kollner TG, Gershenzon J, Degenhardt J (2002) The maize gene terpene synthase 1 encodes a sesquiterpene synthase catalyzing the formation of (*E*)-β-farnesene, (*E*)-nerolidol, and (*E*,*E*)-farnesol after herbivore damage. Plant Physiol **130:** 2049–2060

Seybold SJ, Bohlmann J, Raffa KF (2000) Biosynthesis of coniferophagous bark beetle pheromones and conifer isoprenoids: Evolutionary perspective and synthesis. Can Entomol **132:** 697–753

Shen B, Zheng Z, Dooner HK (2000) A maize sesquiterpene cyclase gene induced by insect herbivory and volicitin: characterization of wild-type and mutant alleles. Proc Natl Acad Sci USA **97:** 14807–14812

Starks CM, Back KW, Chappell J, Noel JP (1997) Structural basis for cyclic terpene biosynthesis by tobacco 5- epi-aristolochene synthase. Science **277:** 1815–1820

Steele CL, Crock J, Bohlmann J, Croteau R (1998a) Sesquiterpene synthases from grand fir (*Abies grandis*): comparison of constitutive and wound-induced activities, and cDNA isolation, characterization, and bacterial expression of δ-selinene synthase and γ-humulene synthase. J Biol Chem **273:** 2078–2089

Steele CL, Katoh S, Bohlmann J, Croteau R (1998b) Regulation of oleoresinosis in grand fir (*Abies grandis*): differential transcriptional control of monoterpene, sesquiterpene, and diterpene synthase genes in response to wounding. Plant Physiol **116:** 1497–1504

Stofer Vogel B, Wildung MR, Vogel G, Croteau R (1996) Abietadiene synthase from grand fir (*Abies grandis*): cDNA isolation, characterization, and bacterial expression of a bifunctional diterpene cyclase involved in resin acid biosynthesis. J Biol Chem **271:** 23262–23268

Trapp SC, Croteau R (2001a) Defensive resin biosynthesis in conifers. Annu Rev Plant Physiol Plant Mol Biol **52:** 689–724

Trapp SC, Croteau RB (2001b) Genomic organization of plant terpene synthases and molecular evolutionary implications. Genetics **158:** 811–832

Van Gunsteren WF, Billeter SR, Eising AA, Hünenberger PH, Krüger P, Mark AE, Scott WRP, Tironi IG (1996) Biomolecular Simulation: The GROMOS96 Manual and User Guide. VDF Hochschulverlag, ETH Zürich, Switzerland

Wang SX, Hunter W, Plant A (2000) Isolation and purification of functional total RNA from woody branches and needles of Sitka and white spruce. Biotechniques **28:** 292–296

Wang XQ, Tank D, Sang T (2000) Phylogeny and divergence times in pinaceae: evidence from three genomes. Mol Biol Evol **17:** 773–781

Whittington DA, Wise ML, Urbansky M, Coates RM, Croteau RB, Christianson DW (2002) Bornyl diphosphate synthase: structure and strategy for carbocation manipulation by a terpenoid cyclase. Proc Natl Acad Sci USA **99:** 15375–15380

Wildung MR, Croteau R (1996) A cDNA clone for taxadiene synthase, the diterpene cyclase that catalyzes the committed step of taxol biosynthesis. J Biol Chem **271:** 9201–9204

Williams DC, McGarvey DJ, Katahira EJ, Croteau R (1998) Truncation of limonene synthase preprotein provides a fully active 'pseudomature' form of this monoterpene cyclase and reveals the function of the amino-terminal arginine pair. Biochemistry **37:** 12213–12220

Wise ML, Croteau R (1999) Monoterpene biosynthesis. *In* DE Cane, ed, Comprehensive Natural Products Chemistry: Isoprenoids, Including Carotenoids and Steroids, Vol 2. Pergamon Press, Oxford, pp 97–154

Zeng G (1998) Sticky-end PCR: new method for subcloning. Biotechniques **25:** 206–208